

**DATA CLUSTERING MENGGUNAKAN  
METODOLOGI CRISP-DM UNTUK PENGENALAN  
POLA PROPORSI PELAKSANAAN TRIDHARMA**

**Tesis**

**untuk memenuhi sebagian persyaratan  
mencapai derajat Sarjana S-2  
Program Studi Magister Sistem Informasi**



**Irwan Budiman**

**24010410400031**

**PROGRAM PASCA SARJANA  
UNIVERSITAS DIPONEGORO  
SEMARANG  
2012**

## ABSTRAK

Kualitas sumber daya manusia dosen dapat tercermin dari produktivitas dan kualitas pelaksanaan tridharma (pendidikan, penelitian, pengabdian pada masyarakat dan kegiatan bidang pendukung). Beban Kerja Dosen dan Evaluasi Pelaksanaan Tridharma Perguruan Tinggi (BKD dan EPT-PT) bertujuan untuk menjamin pelaksanaan tugas dosen berjalan sesuai dengan kriteria yang telah ditetapkan dalam peraturan perundang-undangan. Data *clustering* pelaksanaan tridharma diperlukan agar mendapatkan suatu pengetahuan (*knowledge discovery*) tentang pola (*pattern recognition*) pelaksanaan tridharma pada perguruan tinggi. Clustering sebagai salah satu teknik data mining harus terukur, dapat dipercaya dan memenuhi suatu standar yang telah disepakati. CRISP-DM adalah standarisasi *data mining* yang digunakan pada penelitian ini. Hasil data clustering menemukan pola proporsi pelaksanaan tridharma menjadi 3 cluster yang mewakili pola: dosen profesional, dosen manajer dan dosen pengajar.

Kata Kunci: *Clustering*, CRISP-DM, K-Means, Tridharma

## ABSTRACT

Quality of human resources faculty can be reflected from the implementation of productivity and quality Tridharma (education, research, community service and supporting field activities). Lecturer Workload and Evaluation of Higher Education Tridharma (BKD and the EPT-PT) aims to ensure the implementation of the faculty task runs according to the criteria set out in legislation. *Data clustering* Tridharma implementation is needed to get some knowledge of the pattern of Tridharma implementation at college. Clustering as a data mining technique should be scalable, reliable and meet an agreed standard. CRISP-DM is the standardization of data mining is used in this study. The results of data clustering found the pattern of proportion of Tridharma into 3 clusters representing patterns: professionals, managers and teachers.

Keywords: Clustering, CRISP-DM, K-Means, Tridharma

# **BAB I**

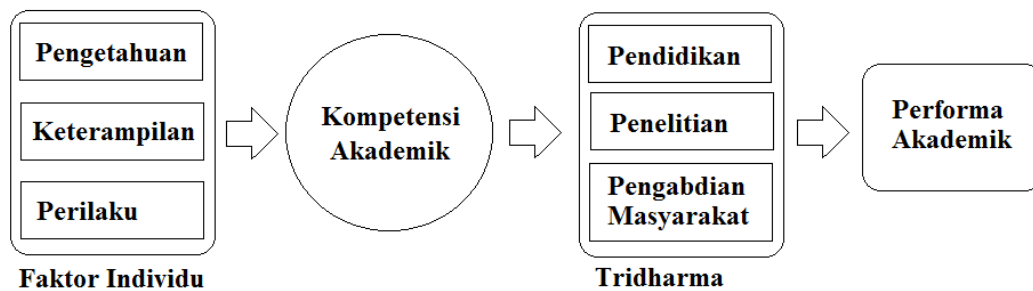
## **PENDAHULUAN**

### **1.1 Latar Belakang**

Dosen adalah salah satu komponen esensial dalam suatu sistem pendidikan di perguruan tinggi. Peran, tugas, dan tanggungjawab dosen sangat penting dalam mewujudkan tujuan pendidikan nasional, yaitu mencerdaskan kehidupan bangsa, meningkatkan kualitas manusia Indonesia, yang meliputi kualitas iman/takwa, akhlak mulia, dan penguasaan ilmu pengetahuan, teknologi, dan seni, serta mewujudkan masyarakat Indonesia yang maju, adil, makmur, dan beradab (Tim Dikti, 2010).

Sebagaimana diamanatkan dalam UU Nomor 14 Tahun 2005 tentang Guru dan Dosen, dosen dinyatakan sebagai pendidik profesional dan ilmuwan dengan tugas utama mentransformasikan, mengembangkan, dan menyebarluaskan ilmu pengetahuan, teknologi, dan seni melalui pendidikan, penelitian, dan pengabdian kepada masyarakat (Bab 1 Pasal 1 ayat 2). Tugas utama dosen adalah melaksanakan tridharma perguruan tinggi dengan beban kerja paling sedikit sepadan dengan 12 (dua belas) sks dan paling banyak 16 (enam belas) sks pada setiap semester sesuai dengan kualifikasi akademik.

Kualitas sumber daya manusia dosen dapat tercermin dari produktivitas dan kualitas pelaksanaan tridharma (pendidikan, penelitian, pengabdian pada masyarakat dan kegiatan pendukung lainnya). Direktorat Jenderal Pendidikan Tinggi (DIKTI) telah memberlakukan program Beban Kerja Dosen dan Evaluasi Pelaksanaan Tridharma Perguruan Tinggi (BKD dan EPT-PT) untuk menjamin pelaksanaan tugas dosen berjalan sesuai dengan kriteria yang telah ditetapkan dalam peraturan perundang-undangan.



Gambar 1.1 Faktor Kompetensi Akademik

Kompetensi tenaga pendidik, khususnya dosen, diartikan sebagai seperangkat pengetahuan, keterampilan dan perilaku yang harus dimiliki, dihayati, dikuasai dan diwujudkan oleh dosen dalam melaksanakan tugas profesionalnya. Kompetensi tersebut meliputi kompetensi pedagogik, kompetensi kepribadian, kompetensi sosial dan kompetensi profesional (Tim Dikti, 2010).

Setiap dosen dapat berbeda pola proporsi pelaksanaan tridharma antara satu dengan yang lainnya. Ada dosen yang proporsi tridharmanya lebih banyak pada bidang pendidikan sehingga bidang penelitian dan pengabdian masyarakatnya menjadi sedikit. Sebagian ada yang fokus pada bidang penelitian sementara bidang pendidikan dan pengabdian masyarakat terabaikan. Ada pula dosen yang proporsi antar bidang tridharmanya relatif seimbang satu sama lainnya.

*Data clustering* pelaksanaan tridharma diperlukan agar mendapatkan suatu pengetahuan (*knowledge discovery*) tentang pola (*pattern recognition*) pelaksanaan tridharma pada perguruan tinggi. Hal ini diperlukan untuk mengontrol keseimbangan kegiatan tridharma dosen dan kemudian oleh lembaga dapat membuat kebijakan yang tepat dan terarah sesuai dengan kondisi dan kebutuhan perguruan tinggi dalam mencapai visi dan misinya.



Gambar 1.2 Ide dasar *data mining* pada perguruan tinggi  
(Delavari, 2008)

Pertumbuhan data pada perguruan tinggi semakin pesat setiap tahunnya. Sehingga menjadi permasalahan tersendiri dalam hal pengelolaannya. Sebuah fakta yang tampak pada institusi perguruan tinggi adalah bahwa pesatnya peningkatan jumlah edukasional data, tidak dibarengi dengan pemanfaatan data tersebut secara maksimal untuk peningkatan manajemen (Delavari, 2008, h.31). Proses tradisional jika didukung teknik *data mining* dapat membantu untuk menemukan pola, struktur dan pengetahuan yang berharga, yang dapat dimanfaatkan oleh instansi terkait sebagai modal untuk peningkatan manajemen khususnya manajemen pengambilan keputusan (**Gambar 1.2**).

*Clustering* merupakan teknik *data mining* yang berfungsi untuk mengelompokkan data berdasarkan kemiripan data pada suatu kelompok dan meminimalkan kemiripan pada kelompok lain. Sehingga dapat diketahui pola yang tersembunyi dan menemukan pengetahuan tentang tipologi pola proporsi pelaksanaan tridharma dosen.

Berdasarkan uraian tersebut di atas, maka penelitian diberi judul “*Data Clustering Menggunakan Metodologi CRISP-DM untuk Pengenalan Pola Proporsi Pelaksanaan Tridharma*”.

### 70.1. Perumusan Masalah

Berdasarkan latar belakang masalah yang diuraikan diatas maka perumusan masalah pada penelitian ini adalah bagaimana mendeskripsikan pola (*pattern recognition*) proporsi pelaksanaan tridharma dengan tehnik *clustering* menggunakan metodologi CRISP-DM sekaligus mengelompokkan dosen berdasarkan proporsi pelaksanaan tridharma.

## 70.2. Batasan Masalah

Batasan masalah dari penelitian ini adalah:

- 1) Database yang diolah adalah database Beban Kerja Dosen dan Evaluasi Pelaksanaan Tridharma pada Universitas Lambung Mangkurat Banjarmasin
- 2) Metodologi *data mining* yang digunakan adalah CRISP-DM
- 3) Algoritma *clustering* yang digunakan adalah algoritma K-means
- 4) Pemodelan *data mining* menggunakan program aplikasi Microsoft Access 2007.

## 70.3. Keaslian Penelitian

Penelitian aplikasi *data mining* pada perguruan tinggi telah banyak dilakukan sebelumnya seperti yang terangkum pada “*Data Mining Applications in Higher Learning Institutions*” (Delavari, 2008). Namun diantara penelitian-penelitian tersebut sedikit sekali yang menjadikan entitas dosen sebagai objek penelitian *data mining*. Dan pula tidak satu pun yang mengangkat tema *clustering* pada entitas dosen. Fakta ini bisa dilihat pada Tabel 2.1 dan Tabel 2.2. Hal ini disebabkan antara lain ketiadaan atribut (data) untuk pemahaman lebih jauh tentang kinerja dosen pada database perguruan tinggi (Delavari, 2008). Penelitian ini pun menggunakan database Beban Kerja Dosen sebagai sumber data karena data tidak tersedia pada database perguruan tinggi.

Pengorganisasian data ke dalam kelompok atau dalam teknik *data mining* disebut sebagai *clustering* merupakan dasar pemahaman dan penggalian pengetahuan. Misalnya, skema umum dari klasifikasi ilmiah menempatkan organisme ke dalam sistem taksa peringkat: domain, kingdom, filum, kelas, dan lain-lain (Jain, 2009). *Clustering* dosen perlu dilakukan sebagai fundamental pemahaman dan penggalian pengetahuan tentang objek dosen sendiri.

Metodologi *data mining* ada beberapa diantaranya CRISP DM pada jurnal “*The CRISP-DM: The New Blueprint for data Mining*” (Shearer, 2010). Dan metodologi ini digunakan pada “*CRISP-DM Model Applied for Knowledge Discovery in Speech Disorder Therapy Area*” (Danubianu, 2009). Artikel terkait

CRISP-DM umumnya bersifat generic dan tidak spesifik menerapkan CRISP-DM pada aplikasi *clustering*, khususnya *clustering* domain dosen berdasarkan pola proporsi pelaksanaan tridharma perguruan tinggi pada database Beban Kerja Dosen yang mulai diberlakukan Dikti sejak 2010 lalu.

Penggunaan K-means sebagai algoritma *clustering* digunakan karena terbukti cukup handal sebagaimana yang diungkapkan pada penelitian “*Data clustering: 50 years beyond K-means*” (Jain, 2009) dan “*Top 10 Algorithm in Data Mining*” (Wu, 2008). Penelitian yang menggunakan K-means antara lain “*Aplikasi K-means untuk Pengelompokan Mahasiswa berdasarkan Nilai BMI dan Ukuran Kerangka*” (Rismawan, 2008). Namun penelitian tersebut tidak menggunakan tahapan CRISP-DM sebagai metodologi.

#### **70.4. Tujuan Penelitian**

Tujuan akhir dari penelitian ini adalah sebagai berikut:

- 1) Menggali pengetahuan (*discovering knowledge*) tentang pola (*pattern*) proporsi pelaksanaan tridharma oleh dosen sehingga diketahui tingkat produktivitas, komposisi serta keseimbangan dalam pelaksanaan tridharma
- 2) Memberi nilai tambah dengan menambah unsur *data mining* terhadap program Beban Kerja Dosen dan Evaluasi Pelaksanaan Tridharma Perguruan Tinggi.

#### **70.5. Manfaat Penelitian**

Manfaat yang diperoleh saat penelitian ini berhasil dilakukan adalah:

- 1) Dapat digunakan untuk menemukan pengetahuan eksplisit baru tentang pola pelaksanaan tridharma oleh dosen
- 2) Dapat digunakan sebagai bahan untuk evaluasi keseimbangan tridharma dosen dalam pelaksanaan tridharma perguruan tinggi
- 3) Dapat digunakan untuk mendukung keputusan dalam perencanaan program strategis berkaitan pembinaan dan program penjaminan mutu perguruan tinggi dan dosen untuk meningkatkan kinerjanya.



## BAB II

### TINJAUAN PUSTAKA

#### 71.1. Tinjauan Pustaka

##### 71.1.1. Penerapan *Data Mining* pada Institusi Perguruan Tinggi.

Salah satu tantangan terbesar yang dihadapi perguruan tinggi saat ini adalah bagaimana meningkatkan kualitas manajemen pengambilan keputusan. Proses pengambilan keputusan menjadi semakin kompleks seiring meningkatnya jumlah data edukasi. Perguruan tinggi dituntut untuk menemukan cara agar manajemen sistem pengambilan keputusan dapat dilakukan dengan lebih efisien. Sebuah fakta yang tampak pada institusi perguruan tinggi adalah bahwa pesatnya peningkatan jumlah edukasional data, tidak dibarengi dengan pemanfaatan data tersebut secara maksimal untuk peningkatan manajemen (Delavari, 2008).

John Naisbitt pada 1982 telah meramalkan berbagai fenomena yang dunia hadapi saat ini dalam terminologi *megatrends*, yaitu tentang berbagai perubahan di bidang sosial, politik, ekonomi dan teknologi. Trend pertama dari sepuluh, Naisbitt mengungkapkan bahwa akan ada perubahan sosial dari era industri menuju era informasi. Era dimana informasi dapat dengan mudah didapat dengan bantuan satelit dan perkembangan teknologi seperti yang kita hadapi dewasa ini. Namun ironisnya dengan *trend* ini Naisbitt juga mengatakan bahwa “Masyarakat memiliki kekayaan akan informasi, namun miskin akan pengetahuan” (Naisbitt, 1982).

Data dan informasi adalah sumber daya yang signifikan dalam institusi perguruan tinggi. Faktanya informasi yang melimpah ruah tersebut berfungsi masih sebatas dokumentasi atau catatan historis semata. Informasi tersebut tidak maksimal memberikan pengetahuan dan kontribusi terhadap peningkatan pada manajemen sistem pengambilan keputusan. Martin Pigott, kepala operasi layanan T-Mobile mengatakan “Memiliki terlalu banyak informasi jauh lebih buruk daripada memiliki terlalu sedikit informasi” (Baxter, 2005).

Institusi perguruan tinggi mencari cara mengelola dan mendukung prosedur pengambilan keputusan untuk membantu mereka dalam mengatur

strategi dan rencana untuk manajemen yang lebih baik. Salah satu cara adalah dengan menggali pengetahuan (*knowledge discovery*) dan pengenalan pola (*pattern recognition*) terkait dengan proses pendidikan dan entitasnya. Pengetahuan dan pengenalan pola ini dapat diekstraksi dari data historis dan operasional yang berada di *database* organisasi pendidikan dengan menggunakan teknik *data mining*.

*Data mining* sebagai salah satu cabang ilmu yang relatif baru mempunyai potensi pengembangan yang sangat besar dan diprediksi akan menjadi salah satu yang paling revolusioner pada dekade ini (Larose, 2006). Teknik *data mining* merupakan sebuah proses ekstraksi informasi untuk menggali pengetahuan (*knowledge discovery*) dan menemukan pola (*pattern recognition*) pada tumpukan data dalam *database* yang biasanya berskala besar. Fungsi-fungsi dalam *data mining* antara lain: fungsi deskripsi, fungsi estimasi, fungsi Prediksi, fungsi Klasifikasi, fungsi *Clustering* dan fungsi asosiasi (Larose, 2005).

Berdasarkan jurnal "*Data Mining Applications in Higher Learning Institutions*" (Delavari, 2008) diungkapkan beberapa contoh penerapan *data mining* pada instansi perguruan tinggi. Antara lain untuk evaluasi *students assessment*, *lecturer assessment*, *course assessment*, *industrial training assessment*, *students registration evaluation*.

Jurnal Delavari ini membahas kemampuan *data mining* di lembaga perguruan tinggi dan mengusulkan beberapa penerapannya. Pada **Tabel 2.1** disajikan *summary* penerapan *data mining* yang dilakukan sebelumnya, **Tabel 2.2** disajikan usulan penerapan *data mining* pada perguruan tinggi.

Tabel 2.1 Summary *Data Mining* pada Perguruan Tinggi

ID	Performance Indicator Type	Objective	Data Mining Method	Explicit Knowledge	Educational Main Process	Educational Sub-Process
1	Input indicator	Predicting alumni pledge	Prediction	The pattern of previous graduates contributing to the university activities	Planning	Alumni activity planning
2	Process indicator	Use of Data Mining in CRCT scores	Prediction	The patterns of previous student test score associated with their gender, race, attendance and so on	Evaluation	Course assessment
3		Creating meaningful learning outcome typologies	Cluster analysis	The patterns of previous student's learning outcome		Student assessment
4		Use data mining techniques to develop institutional typologies	Cluster analysis	The pattern of various groups of students		
5		Academic planning and intervention transfer prediction	Prediction	The success patterns of previous students who had previously transferred subjects		
6		Predicting and clustering persisters and non-persisters	Prediction Clustering	The patterns of previous similar student who were persistent or non persistent		
7		Predicting student performance	Classification	Classified pattern of previous students based on their final grade		
8	Output indicator	Improving quality of graduate student by data mining	Association Classification	Characteristic patterns of previous students who took a particular major and The patterns of previous students which were likely to be good in a given major	Counseling	Student major counseling

(Sumber: Delavari, 2008)

Berdasarkan **Tabel 2.1**, fokus utama dari banyak peneliti adalah pada indikator proses (*process indicator*) dari domain pendidikan. Hal ini juga dapat diamati bahwa ada banyak perhatian pada proses evaluasi. Penekanan juga ada pada sub proses evaluasi pada enititas mahasiswa. Namun hasil dari studi ini menunjukkan bahwa masih ada banyak daerah dalam domain pendidikan yang belum ditambang dan luput dari pandangan peneliti *data mining*. Jurnal Delavari mengidentifikasi daerah-daerah dalam domain pendidikan, dan menjadi usulan untuk penambangan data berikutnya. Jurnal ini memaparkan area-area yang memungkinkan untuk dilakukannya penelitian *data mining* pada lembaga pendidikan tinggi.

Tabel 2.2 Usulan Penerapan *Data Mining* pada Perguruan Tinggi

Main Process	Sub-Process	Explicit Knowledge	Data Mining Method
Evaluation	Student assessment	<ul style="list-style-type: none"> <li>• The success patterns of previous students who previously had transferred subjects</li> </ul>	<ul style="list-style-type: none"> <li>• Prediction</li> </ul>
		<ul style="list-style-type: none"> <li>• The patterns of previous students who were likely to be good in a given major</li> <li>• The patterns and relationship of various factors affecting the student test score</li> <li>• Prediction of the likelihood of success</li> </ul>	<ul style="list-style-type: none"> <li>• Prediction</li> </ul>
		<ul style="list-style-type: none"> <li>• The success patterns of previous similar students</li> <li>• Prediction of likelihood of persistence</li> </ul>	<ul style="list-style-type: none"> <li>• Prediction</li> <li>• Clustering</li> </ul>
		<ul style="list-style-type: none"> <li>• The patterns of previous successful and unsuccessful graduates</li> </ul>	<ul style="list-style-type: none"> <li>• Prediction,</li> <li>• Clustering</li> </ul>
		<ul style="list-style-type: none"> <li>• The patterns of previous students who planned to drop subject</li> </ul>	<ul style="list-style-type: none"> <li>• Prediction</li> </ul>
		<ul style="list-style-type: none"> <li>• The patterns of previous students who planned for resource allocation</li> </ul>	<ul style="list-style-type: none"> <li>• Prediction</li> </ul>
		<ul style="list-style-type: none"> <li>• The patterns of previous male and female students in test score</li> </ul>	<ul style="list-style-type: none"> <li>• Association</li> </ul>
		<ul style="list-style-type: none"> <li>• The patterns of previous student's learning outcome</li> </ul>	<ul style="list-style-type: none"> <li>• Prediction,</li> <li>• Clustering</li> </ul>
		<ul style="list-style-type: none"> <li>• The patterns of previous students attendance in accordance with test score</li> </ul>	<ul style="list-style-type: none"> <li>• Association</li> </ul>
		<ul style="list-style-type: none"> <li>• Association of student health information and test score</li> </ul>	<ul style="list-style-type: none"> <li>• Association</li> </ul>
	Lecturer assessment	<ul style="list-style-type: none"> <li>• The characteristic patterns of previous lecturers which were more effective than others</li> </ul>	<ul style="list-style-type: none"> <li>• Prediction,</li> <li>• Classification</li> </ul>
		<ul style="list-style-type: none"> <li>• Association between lecturer training and student test score</li> </ul>	<ul style="list-style-type: none"> <li>• Association</li> </ul>
	Course assessment	<ul style="list-style-type: none"> <li>• Cluster of most cost-effective courses to be offered together</li> </ul>	<ul style="list-style-type: none"> <li>• Clustering</li> </ul>
		<ul style="list-style-type: none"> <li>• The patterns of courses who offered previously to different type of students</li> </ul>	<ul style="list-style-type: none"> <li>• Classification</li> <li>• Association</li> </ul>
		<ul style="list-style-type: none"> <li>• Prediction of factors most affected in test score in various courses</li> </ul>	<ul style="list-style-type: none"> <li>• Prediction</li> </ul>

		<ul style="list-style-type: none"> <li>• The patterns of programs (courses) which produce greatest return and investment in terms of student learning in coming year</li> </ul>	<ul style="list-style-type: none"> <li>• Prediction</li> </ul>
	Industrial training assessment	<ul style="list-style-type: none"> <li>• The patterns of previous training course for different type of student</li> </ul>	<ul style="list-style-type: none"> <li>• Classification</li> <li>• Association</li> </ul>
	Student registration evaluation	<ul style="list-style-type: none"> <li>• The success patterns of those students who successfully enrolled to the university</li> </ul>	<ul style="list-style-type: none"> <li>• Prediction</li> </ul>
Planning	Course Planning	<ul style="list-style-type: none"> <li>• Classification of courses to most appropriate time</li> <li>• Success patterns of courses which were taken together</li> </ul>	<ul style="list-style-type: none"> <li>• Classification</li> <li>• Clustering</li> </ul>
	Academic planning	<ul style="list-style-type: none"> <li>• The patterns of previous discipline problems in academic planning</li> </ul>	<ul style="list-style-type: none"> <li>• Prediction</li> </ul>
	Lecturer time table planning	<ul style="list-style-type: none"> <li>• The patterns of previous lecturer's class time table</li> <li>• Prediction of lecturer time table for coming year</li> </ul>	<ul style="list-style-type: none"> <li>• Prediction</li> </ul>
	Alumni activities planning	<ul style="list-style-type: none"> <li>• The pattern of previous graduates contributing in university activities</li> <li>• Prediction of the likelihood of alumni who continued studies</li> <li>• Prediction of the likelihood of alumni who find suitable job</li> </ul>	<ul style="list-style-type: none"> <li>• Prediction</li> </ul>
Registration	Student course registration	<ul style="list-style-type: none"> <li>• The patterns of previous students who take various subjects</li> <li>• Association of student to the most appropriate subject</li> </ul>	<ul style="list-style-type: none"> <li>• Prediction</li> <li>• Association</li> </ul>
Counseling	Student behavioral consulting	<ul style="list-style-type: none"> <li>• The patterns of previous students behavior in an academic environment</li> </ul>	<ul style="list-style-type: none"> <li>• Clustering</li> </ul>
	Major selection consulting	<ul style="list-style-type: none"> <li>• The characteristic patterns of previous students who took particular major</li> </ul>	<ul style="list-style-type: none"> <li>• Classification</li> <li>• Association</li> </ul>
	Course selection consulting	<ul style="list-style-type: none"> <li>• Classification of student to various elective subject</li> <li>• Classification of student to various courses</li> </ul>	<ul style="list-style-type: none"> <li>• Classification</li> </ul>
	Program Counseling	<ul style="list-style-type: none"> <li>• The patterns of previous student who were good in a given program</li> </ul>	<ul style="list-style-type: none"> <li>• Association</li> <li>• Classification</li> </ul>

Examination	Student examination	<ul style="list-style-type: none"> <li>• Association between exam level &amp; student mark</li> <li>• Association between exam level and lecturer class performance</li> </ul>	• Association
Performance	Student performance	<ul style="list-style-type: none"> <li>• Association between student performance and lecturer satisfaction</li> <li>• Association between student course mark and time and venue of classes</li> </ul>	• Association
	Lecturer performance	<ul style="list-style-type: none"> <li>• Association between lecturer who cancel the class frequently and student test score</li> <li>• Association between lecturer background and time and his/her performance</li> </ul>	• Association
Marketing	University advertising	• The characteristic patterns of previous international lecturer and student which attract to the universities	• Prediction
		• The patterns of previous local student & lecturer who resign or terminate from local universities	• Prediction

(Sumber: Delavari, 2008)

**Tabel 2.2** menyajikan usulan *data mining* pada perguruan tinggi. Kolom pertama dari tabel sesuai adalah proses utama yang biasanya terjadi dalam sebuah lembaga pendidikan tinggi. Teridentifikasi tujuh proses utama dalam sistem perguruan tinggi tinggi, yaitu "evaluasi", "perencanaan", "pendaftaran", "konsultasi", "pemasaran", "kinerja" dan "pengawasan". Setiap proses dikategorikan ke dalam beberapa detail sub-proses, yang disajikan dalam kolom dua. Kolom ketiga dari panduan ini memperkenalkan set pengetahuan dan informasi, yang sebenarnya adalah serangkaian pola ekstraksi dan struktur dari data dengan menggunakan teknik dan algoritma *data mining*. Ini berarti bahwa menerapkan teknik *data mining* pada set data, serangkaian pengetahuan yang bermakna dapat ditemukan. Kolom terakhir dari pedoman menggambarkan metode *data mining* yang paling tepat untuk mengekstraksi pengetahuan yang bermanfaat.

Usulan pada **Tabel 2.2** hanyalah sedikit dari sekian banyak area pada domain perguruan tinggi yang dapat diterapkan untuk proyek *data mining*. Namun usulan tersebut dapat menjadi tuntunan dan menjadi peta penelitian *data mining* di instansi perguruan tinggi. Lebih jauh bahkan *data mining* dapat digunakan untuk mengelompokkan mahasiswa berdasarkan status gizi dan ukuran kerangkanya, mengacu pada hasil penelitian “**Aplikasi K-means untuk Pengelompokan Mahasiswa berdasarkan Nilai BMI dan Ukuran Kerangka**” (Rismawan, 2008).

Model *data mining* pada perguruan tinggi diklasifikasikan dalam dua kategori utama, yaitu model deskriptif dan model prediktif. Model deskriptif menggambarkan kumpulan data dalam cara yang ringkas dan dirangkum dan menyajikan sifat umum yang menarik dari data. Ini menjelaskan pola dalam data yang ada, yang dapat digunakan untuk memandu menuju proses pengambilan keputusan. Model prediktif memprediksi perilaku yang didasarkan pada data historis dan menggunakan data dengan hasil yang diketahui untuk membangun sebuah model yang dapat kemudian digunakan untuk memprediksi secara eksplisit nilai untuk data yang berbeda (Delavari, 2008).

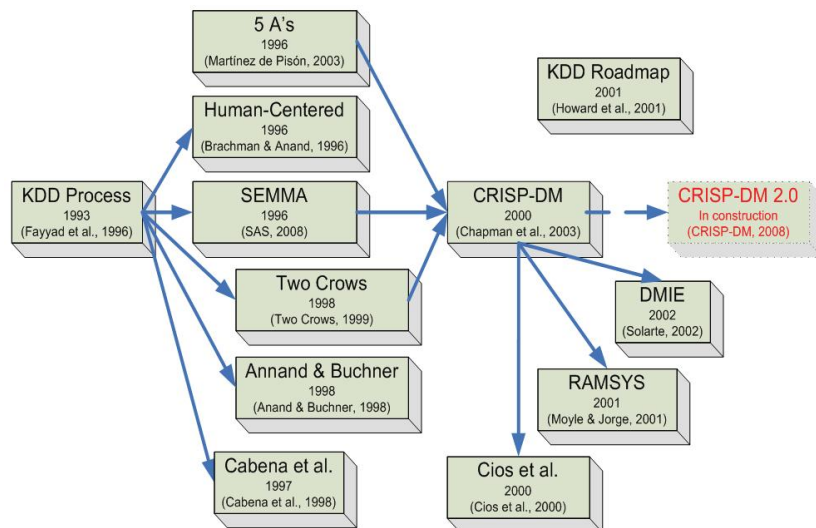
Proses *data mining* harus terukur, dapat dipercaya dan memenuhi suatu standar yang telah disepakati. CRISP-DM adalah standarisasi metodologi *data mining* yang disusun oleh tiga penggagas *data mining market*, yaitu Daimler Chrysler (Daimler-Benz), SPSS (ISL), NCR (Shearer, 2000). CRISP-DM dikembangkan pada berbagai *workshops* antara 1997-1999. Jurnal “**Data Mining Applications in Higher Learning Institutions**” merekomendasikan penggunaan metodologi CRISP-DM sebagai proses model *data mining* (Delavari, 2008).

### 71.1.2. Perkembangan Metodologi dan Model *Data Mining*

Dalam beberapa literatur, penggunaan terminologi ‘model proses’, ‘*life cycle*’, ‘metodologi’ dan ‘*framework*’ dipakai untuk menunjuk pada hal yang sama. Satu model proses atau metodologi atau *framework* adalah satu set dari pekerjaan yang dilaksanakan untuk mengembangkan satu unsur / pekerjaan tertentu. Tujuan dari satu model proses adalah untuk membuat proses agar dapat diulang, dikendalikan dan terukur (Marban, 2009).

Salah satu tuntutan dari *data mining* ketika diterapkan pada data berskala besar adalah diperlukan metodologi sistematis tidak hanya ketika melakukan analisa saja tetapi juga ketika mempersiapkan data dan juga melakukan interpretasi dari hasilnya sehingga dapat menjadi aksi ataupun keputusan yang bermanfaat (Kusnawi, 2007).

Metodologi sistematis *data mining* telah mengalami perkembangan seiring berjalannya waktu. Berawal dengan ditetapkan metodologi KDD yang merupakan singkatan dari *Knowledge Discovery in Database* pada awal tahun 1990-an. Kemudian beberapa metodologi *data mining* lain menyusul sesudahnya. Beberapa yang cukup populer adalah SEMMA dan CRISP-DM. Pada **Gambar 2.1** dapat dilihat tahapan perkembangan evolusi metodologi *data mining* dari KDD hingga CRISP-DM.



Gambar 2.1 Evolusi Metodologi *Data Mining* (Marban, 2009)

Istilah KDD (*Knowledge Discovery in Database*) mulai dikenal pada era awal 1990-an. Fayyad pada 1996 mengagas proses model KDD dan menetapkan langkah untuk proyek DM. Kemudian pada perkembangannya banyak dimodifikasi seperti Brachman & Anand, SEMMA, Two Crows, Anand & Buchner dan Cabena.



Pada lain sisi 5A's adalah model proses yang mengusulkan tugas-tugas proyek DM dan merupakan salah satu pelopor CRISP-DM. CRISP-DM memodifikasi 5A's, SEMMA dan Two Crows. Pada perkembangannya CRISP-DM telah menjadi standard *de facto* untuk proses *data mining*. Perusahaan-perusahaan yang mengembangkan CRISP-DM telah bergabung untuk mengembangkan CRISP-DM 2.0. Namun hingga kini CRISP-DM 2.0 belum diterbitkan.

CRISP-DM menginspirasi CIOs yang lebih fokus pada kebutuhan riset akademik, RAMSYS (*RAPid collaborative data Mining SYStem*) dan DMIE (*Data Mining for Industrial Engineering*). Metodologi ini didasarkan pada CRISP-DM dan mempertahankan fase yang sama secara generik.

Roadmap KDD adalah metodologi DM digunakan dalam menjelaskan proses yang tersedia dan algoritma dan menggabungkan pengalaman berasal dari proyek-proyek komersial yang berhasil diselesaikan.

Meski pun pada perkembangannya metodologi *data mining* telah mengalami modifikasi dan pengembangan, namun secara garis besar, tahapannya dapat digolongkan menjadi 3 tahapan utama, yaitu:

- 1) *pre-processing*, tahap persiapan
- 2) *data mining*, tahap penambangan data
- 3) *post-processing*, tahap setelah *data mining* (Liu, 2007).

Pada **Tabel 2.3** dapat diamati bahwa meski pun berbeda detail tahapan model prosesnya, namun secara garis besar ketiga tahapan besar tersebut adalah benang merah dari metodologi *data mining*.

**Tabel 2.3** Perbandingan Model dan Metodologi *Data Mining*

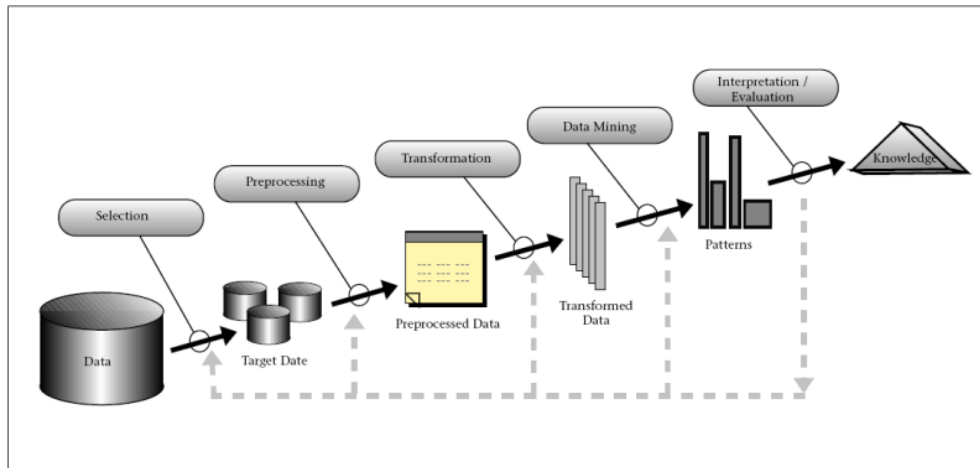
Model	Fayyad et al.	Cabena et al.	Anand & Buchner	CRISP-DM	Cios et al.
No of steps	9	5	8	6	6
Steps	Developing and Understanding of the Application Domain	Business Objectives Determination	Human Resource Identification Problem Specification	Business Understanding	Understanding the Data
	Creating a Target Data Set	Data Preparation	Data Prospecting	Data Understanding	Understanding the Data
	Data Cleaning and Pre-processing		Domain Knowledge Elicitation		
	Data Reduction and Projection		Methodology Identification	Data Preparation	Preparation of the data
	Choosing the DM Task		Data Pre-processing		
	Choosing the DM Algorithm				
	DM	DM	Pattern Discovery	Modeling	DM
	Interpreting Mined Patterns	Domain Knowledge Elicitation	Knowledge Post-processing	Evaluation	Evaluation of the Discovered Knowledge
	Consolidating Discovered Knowledge	Assimilation of Knowledge		Deployment	Using the Discovered Knowledge

(Sumber: Marban, 2009)

### 71.1.3. Metodologi KDD

Proses KDD adalah proses menggunakan metode *data mining* untuk mengekstrak pengetahuan apa yang dianggap sesuai dengan spesifikasi ukuran dan batas, menggunakan database bersama dengan *preprocessing* yang diperlukan, pengambilan sampel dan transformasi dari database.

Ada lima tahap seperti **Gambar 2.2**. Seleksi, tahap ini terdiri pada menciptakan menetapkan target data, atau memfokuskan pada subset variabel atau sampel data, di mana penemuan yang akan dilakukan. Pra-pengolahan - pada tahap ini terdiri target pembersihan data dan pra pengolahan untuk memperoleh data yang konsisten. Transformasi - tahap ini terdiri pada transformasi data menggunakan reduksi dimensi atau transformasi metode. *Data Mining* - tahap ini terdiri pada mencari pola yang menarik dalam bentuk representasional tertentu , tergantung pada tujuan. Interpretasi / evaluasi - tahap ini terdiri pada interpretasi dan evaluasi pola ditambang (Azevedo, 2008).



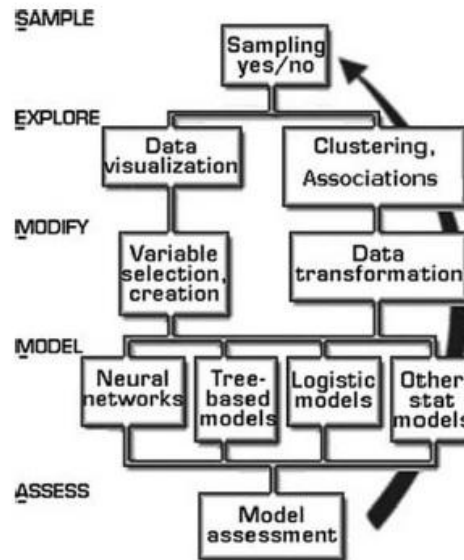
Gambar 2.2 Proses KDD

#### 71.1.4. Metodologi SEMMA

SEMMA (*Sample Explore Modify Model Assess*) mengacu pada proses melakukan sebuah proyek *data mining*. SAS Institute menmbagi siklus dengan 5 tahapan untuk proses *data mining*.

*Sample* - tahap ini terdiri atas sampling data dengan mengekstraksi sebagian data besar ditetapkan cukup besar untuk menampung informasi yang signifikan. *Explore* - tahap ini terdiri pada eksplorasi data dengan mencari tren yang tak terduga dan anomali dalam rangka untuk mendapatkan pemahaman dan ide-ide. *Modify* - tahap modifikasi dari data dengan menciptakan, memilih, dan mengubah variabel untuk fokus proses pemilihan model. *Model* - tahap ini terdiri pada pemodelan data dengan suatu perangkat lunak untuk mencari secara otomatis kombinasi data yang dapat diandalkan untuk memprediksi hasil yang diinginkan. *Assess* - tahap ini terdiri pada menilai data dengan mengevaluasi kegunaan dan keandalan dari temuan dari proses *data mining*.

Proses SEMMA (**Gambar 2.3**) menawarkan yang mudah untuk memahami proses, memungkinkan perkembangan terorganisir dan memadai dan pemeliharaan proyek-proyek *data mining*. Dengan demikian struktur untuk menganugerahkan, penciptaan dan evolusi konsepsinya, membantu untuk menyajikan solusi untuk masalah bisnis serta untuk menemukan tujuan de bisnis DM (Azevedo, 2008).



Gambar 2.3 Proses SEMMA

#### 71.1.5. Metodologi CRISP-DM

CRISP-DM adalah standarisasi *data mining* yang disusun oleh tiga penggagas *data mining market*. Yaitu Daimler Chrysler (Daimler-Benz), SPSS (ISL), NCR. Kemudian dikembangkan pada berbagai *workshops* (antara 1997-1999). Lebih dari 300 organisasi yang berkontribusi dalam proses modelling ini dan akhirnya CRISP-DM 1.0 dipublikasikan pada 1999 (Shearer, 2000).

Proses *data mining* berdasarkan CRISP-DM terdiri dari 6 fase. Yaitu:

*Business Understanding* atau pemahaman domain (penelitian). Pada fase ini dibutuhkan pemahaman tentang substansi dari kegiatan *data mining* yang akan dilakukan, kebutuhan dari perspektif bisnis. Kegiatannya antara lain: menentukan sasaran atau tujuan bisnis, memahami situasi bisnis, menentukan tujuan *data mining* dan membuat perencanaan strategi serta jadwal penelitian.

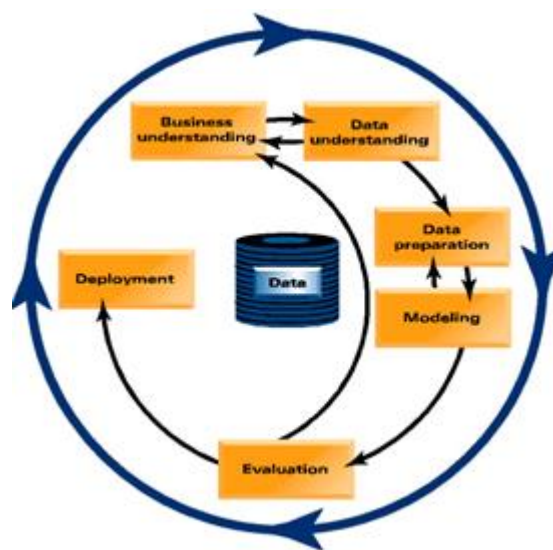
*Data Understanding* atau pemahaman data adalah fase mengumpulkan data awal, mempelajari data untuk bisa mengenal data yang akan dipakai. Fase ini mencoba mengidentifikasi masalah yang berkaitan dengan kualitas data, mendeteksi subset yang menarik dari data untuk membuat hipotesa awal.

*Data preparation* atau persiapan data. Fase ini sering disebut sebagai fase yang padat karya. Aktivitas yang dilakukan antara lain memilih *table* dan *field* yang akan ditransformasikan ke dalam database baru untuk bahan *data mining* (set data mentah).

*Modeling* adalah fase menentukan teknik data mining yang digunakan, menentukan *tools data mining*, teknik data mining, algoritma data mining, menentukan parameter dengan nilai yang optimal.

*Evaluation* adalah fase interpretasi terhadap hasil data mining yang ditunjukkan dalam proses pemodelan pada fase sebelumnya. Evaluasi dilakukan secara mendalam dengan tujuan menyesuaikan model yang didapat agar sesuai dengan sasaran yang ingin dicapai dalam fase pertama.

*Deployment* atau penyebaran adalah fase penyusunan laporan atau presentasi dari pengetahuan yang didapat dari evaluasi pada proses *data mining*.



Gambar 2.4. Proses CRISP-DM (Shearer, 2000)

#### 71.1.6. Perbandingan KDD, SEMMA dan CRISP-DM

Perbandingan KDD (*Knowledge Discovery from Database*), SEMMA (*Sample Explore Modify Model Assess*) dan CRISP-DM (*Cross Industry Standard Process for Data Mining*) dilakukan untuk menganalisa persamaan dan perbedaan substansil di antara metodologi tersebut. Pendekatan KDD and SEMMA secara tegas dinyatakan ekuivalen. *Sample* dapat diidentifikasi dengan *Selection*. *Explore* dapat diidentifikasi sebagai *Pre processing*. *Modify* sebagai *Transformation*. *Model* sebagai DM. *Assess* dapat berarti *Interpretation/Evaluation*. Proses SEMMA terlihat mirip dengan lima tahapan pada KDD.

CRISP-DM tidak seterang SEMMA dalam mengadopsi KDD. Tahap *Business Understanding* dapat diidentifikasi sebagai pemahaman tentang domain bisnis yang akan diterapkan dan apa relevansinya dengan tujuan yang diharapkan oleh pengguna. Tahap

*Deployment* dapat diidentifikasi sebagai tahap konsolidasi mengenai langkah apa yang diambil setelah hasil pemodelan *data mining* diperoleh. Tahap lainnya dapat dikatakan merupakan adaptasi KDD. Tahap *Data Understanding* sebagai kombinasi dari *Selection* dan *Pre processing*. Tahap *Data Preparation* sebagai *Transformation*. Tahap *Modeling* dapat diidentifikasi dengan DM. *Evaluation* dengan *Interpretation/Evaluation* (Azevedo, 2008).

Tabel 2.4 Perbandingan metodologi KDD, SEMMA dan CRISP-DM

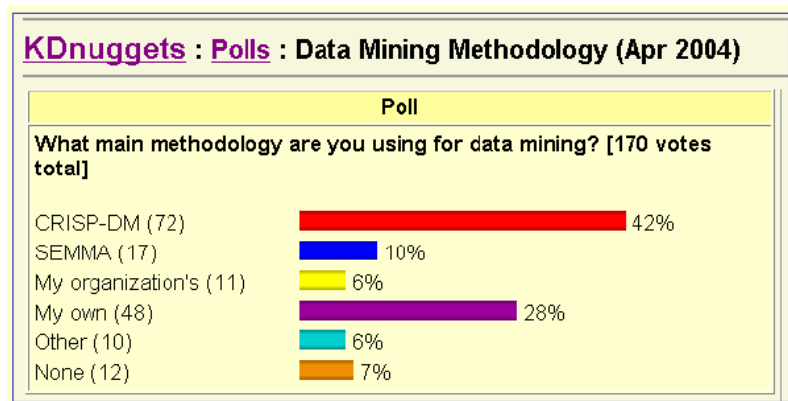
KDD	SEMMA	CRISP-DM
Pre KDD	-----	Business understanding
Selection	Sample	Data Understanding
Pre processing	Explore	
Transformation	Modify	Data preparation
Data mining	Model	Modeling
Interpretation/Evaluation	Assessment	Evaluation
Post KDD	-----	Deployment

(Sumber: Azevedo, 2008).

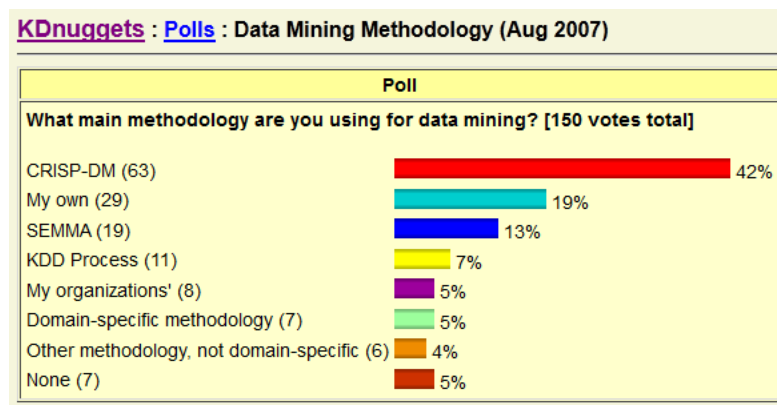
Mempertimbangkan analisa yang disajikan pada **Tabel 2.4** dapat disimpulkan bahwa SEMMA dan CRISP-DM dapat dipandang sebagai satu implementasi hasil adaptasi dari proses KDD. Namun metodologi CRISP-DM lebih kenprehensif dibandingkan SEMMA. Bagaimanapun SEMMA dan CRISP-DM memandu orang-orang untuk mengetahui bagaimana *data mining* dapat diterapkan dalam praktek secara nyata (Azevedo, 2008).

**CRISP-DM menyediakan standar proses baku untuk data mining yang dapat diterapkan ke dalam strategi pemecahan masalah umum pada bisnis atau pada unit penelitian.** CRISP-DM dibandingkan metodologi data mining lain lebih lengkap dan terdokumentasi dengan baik. Setiap fase terstruktur dan terdefinisi dengan jelas sehingga mudah untuk diaplikasikan bahkan bagi pemula sekali pun. (Danubianu, 2009).

Kdnuggets melakukan *survey* penggunaan metodologi *data mining* pada April 2004 dan Agustus 2007. Hasil *survey* menunjukkan bahwa penggunaan CRISP-DM mengungguli metodologi lain termasuk SEMMA dan KDD dengan 42% koresponden memilih CRISP-DM, 13% koresponden memilih SEMMA, dan 7% koresponden memilih KDD (Kdnuggets, 2007).



Gambar 2.5 Kdnuggets *Polls* Tahun 2004



Gambar 2.6 Kdnuggets *Polls* Tahun 2007

#### 71.1.7. Data Clustering Menggunakan Algoritma K-means

Pada artikel yang diterbitkan Elsevier “*Data clustering: 50 years beyond K-means*” (Jain, 2009) diungkapkan bahwa pengorganisasian data ke dalam suatu kluster merupakan suatu model yang paling mendasar untuk pemahaman dan pembelajaran. Analisis kluster adalah studi formal untuk mengelompokkan, atau *clustering* benda-benda sesuai dengan karakteristik yang diukur berdasarkan kemiripan satu sama lain. *Clustering* adalah pengelompokan menggunakan teknik *unsupervised learning* dimana tidak diperlukan pelatihan pada metode tersebut atau dengan kata lain, tidak ada fase *learning* serta tidak menggunakan pelabelan pada setiap kelompok.

Pada jurnal yang sama diungkapkan pula perihal penggunaan K-means sebagai algoritma yang sejak dulu bahkan sampai saat ini cukup handal digunakan untuk proses *clustering*. Dalam perkembangannya K-means juga telah melalui beberapa modifikasi antara lain Fuzzy C-means. Namun jurnal “*Performance Evaluation of K-means and Fuzzy*

*C-Means Clustering Algorithms for Statistical Distributions of Input Data Points*” (Velmurugan, 2010) mengungkapkan bahwa melalui serangkaian percobaan didapatkan kesimpulan bahwa performansi algoritma K-means lebih baik daripada algoritma fuzzy c-means.

Algoritma K-means adalah algoritma yang terbaik dalam algoritma *partitional clustering* dan yang paling sering digunakan diantara algoritma *clustering* lainnya karena kesederhanaan dan efisiensinya (Liu, 2007).

Pada survey yang dipublikasikan Springer “*Top 10 Algorithm in Data Mining*” (Wu, 2008), algoritma K-means ditempatkan pada posisi 2 (dua) sebagai algoritma paling banyak digunakan dalam *data mining* dan menjadi posisi pertama untuk algoritma *clustering*. Urutan *Top* Algoritma-nya adalah sebagai berikut :

- 1) C4.5
- 2) K-means
- 3) SVM (*Support Vector machines*)
- 4) Algoritma Apriori
- 5) EM (*Expectation Maximazation*)
- 6) Algoritma PageRank
- 7) Algoritma AdaBoost
- 8) K-Nearst Neighbor
- 9) Naive Bayes
- 10) Classification and Regression Trees.

#### **71.1.8. Keseimbangan dalam Tridharma**

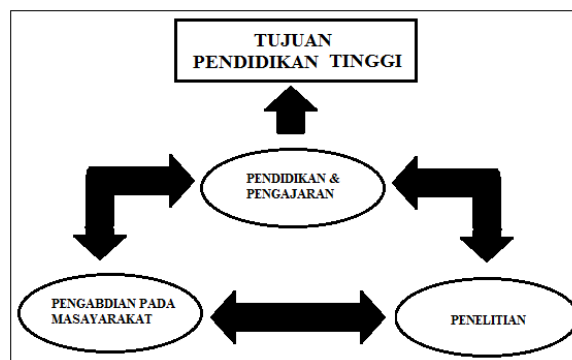
Kualitas sumber daya manusia dosen dapat tercermin dari produktivitas dan kualitas pelaksanaan tridharma. Tridharma, sesuai dengan falsafah yang diamanatkan oleh pemerintah bagi institusi ini yaitu kegiatan dalam bidang Pendidikan, bidang Penelitian dan bidang Pelayanan Masyarakat.

Penyelenggaraan tridharma harus ada dalam koordinasi, mengingat adanya saling interaksi antar tiap kegiatannya. Diinginkan adanya keseimbangan kegiatan melalui pengerahan segenap daya dan potensi yang dimiliki perguruan tinggi yang bersangkutan hingga ke 3 bidang dalam kegiatan Tridharma berikut unsur-unsurnya berkembang sinkron. Kepincangan dalam salah satu tridharma akan menghambat tercapainya tujuan pendidikan tinggi (Lubis, 2004).



Tujuan pendidikan tinggi sebagaimana dalam PP No 60 tahun 1999 adalah menyiapkan peserta didik menjadi anggota masyarakat yang memiliki kemampuan akademik dan/atau profesional yang dapat menerapkan, mengembangkan dan/atau memperkaya khasanah ilmu pengetahuan, teknologi dan/atau kesenian; mengembangkan dan menyebarluaskan ilmu pengetahuan, teknologi dan/atau kesenian serta mengupayakan penggunaannya untuk meningkatkan taraf kehidupan masyarakat dan memperkaya kebudayaan nasional.

Tujuan pendidikan tinggi dapat tercapai jika tridharma dilaksanakan secara proporsional. Tridharma memiliki keterkaitan satu sama lain. Bidang pendidikan memiliki kontribusi terhadap bidang penelitian dalam hal memperkuat dasar teoritik dan penelitian menemukan khasanah baru dalam bidang pendidikan. Bidang penelitian berkontribusi terhadap bidang pengabdian masyarakat dalam hal penerapan hasil penelitian untuk meningkatkan taraf kehidupan masyarakat dan memperkaya kebudayaan. Bidang pengabdian masyarakat dapat menjadi studi kasus dalam bidang pendidikan dan menjadi bahan masukan jika ditemukan masalah atau fenomena baru dalam masyarakat untuk bahan penelitian (Hendrajaya, 2009).



Gambar 2.7 Keterkaitan Tridharma (Lubis, 2004)

Pada kenyataannya keseimbangan dalam pelaksanaan tridharma menjadi permasalahan tersendiri bagi institusi. Pada artikel “Implementasi Tridharma Perguruan Tinggi dalam Mendukung Disiplin Nasional” (Lubis, 2004) memetakan permasalahan yang biasanya dihadapi dosen dan institusi perguruan tinggi pada masing-masing tridharma.

Permasalahan pada bidang pendidikan dan pengajaran:

- 1) Kurikulum Perguruan Tinggi yang berlaku pada saat ini dirasa belum menjawab tantangan global

- 2) Kondisi perangkat keras dan lunak di Perguruan Tinggi yang sudah tidak memadai
- 3) Lemahnya tingkat koordinasi dan keterpaduan antar lembaga yang bertanggung jawab dalam dunia pendidikan
- 4) Pengaruh globalisasi yang tidak dapat dihindarkan
- 5) Belum terciptanya iklim yang kondusif agar gagasan dan pendapat benar-benar didasarkan atas pemikiran rasional yang didukung oleh bukti yang biasa ditinjau kebenarannya
- 6) Masih timpangnya komposisi keahlian antara ilmu sosial dan sains teknologi.

Permasalahan pada bidang penelitian:

- 1) Kemampuan dasar untuk meneliti masih bervariasi, hal ini disebabkan karena jenjang pendidikan yang dicapainya berbeda-beda. Kesempatan untuk mengadakan penelitian sering terhambat oleh tugas-tugas seperti beban tugas mengajar yang masih cukup besar
- 2) Peluang. Keterbatasan memperoleh kesempatan mengajukan proposal yang dapat disetujui, dapat mengurangi dorongan untuk mengajukan proposal penelitian selanjutnya. Tawaran mengajukan proposal penelitian antara penelitian sosial dan non sosial belum berimbang
- 3) Dana. Sumber dana yang ada di perguruan tinggi maupun di DIKTI tersedia terbatas
- 4) Keterbatasan fasilitas, baik sarana maupun prasarana di beberapa perguruan tinggi kurang mendukung kegiatan penelitian.

Permasalahan pada bidang pengabdian masyarakat:

- 1) Pelaksana yang merasa lebih dari masyarakat. Dosen yang melaksanakan pengabdian pada masyarakat sering menganggap masyarakat tersebut bodoh. Akibatnya terjadi pemaksaan pendapat dari dosen kepada masyarakat
- 2) Keragaman budaya yang berakibat terjadinya perbedaan penilaian terhadap sesuatu atau perilaku tertentu dari masyarakat setempat ataupun para pelaksana
- 3) Kekurangtahuan pelaksana dalam berkomunikasi, sehingga dapat menimbulkan kesalah pengertian antara masyarakat dan pelaksana dan dapat berdampak negatif terhadap program, masyarakat maupun pelaksana
- 4) Kemiskinan dan kekurangtahuan masyarakat, sehingga sulit membangkitkan motivasi untuk diajak maju bersama
- 5) Kurangnya koordinasi antar dan dalam institusi pelaksana



Apabila ketercapaian kinerja dosen tersebut telah memenuhi syarat seperti yang dimaksud pada pedoman BKD dan bukti pendukung sesuai dengan laporan yang dibuat maka laporan kinerja dianggap lolos. Bukti pendukung laporan yang telah lolos dikembalikan kepada dosen yang bersangkutan untuk disimpan kembali dan dapat ditunjukkan apabila diperlukan. Kedua asesor menandatangani Format F1 dan meneruskan format F1 kepada Dekan atau yang sederajat untuk mendapatkan pengesahan.

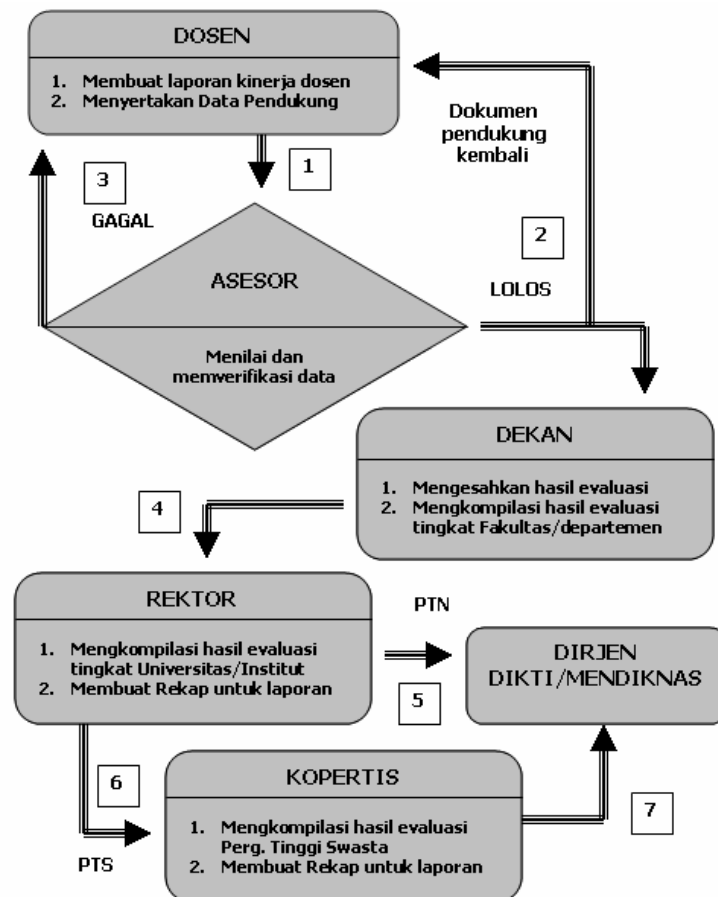
Apabila asesor menyatakan (a) ketercapaian kinerja dosen tidak atau belum memenuhi syarat dan atau (b) bukti pendukung tidak sesuai dengan aktivitas yang dilaporkan maka laporan kinerja dianggap gagal dan dikembalikan kepada dosen yang bersangkutan, untuk diperbaiki. Dalam hal terjadi selisih pendapat antara asesor satu dengan asesor yang lain maka pemimpin perguruan tinggi dapat menunjuk asesor ketiga.

Dekan mengesahkan hasil laporan format F1 dan mengkompilasi semua laporan kinerja dosen yang menjadi tanggungjawabnya. Dekan bertanggung jawab dan berwenang untuk memverifikasi kebenaran laporan yang telah dikoreksi oleh asesor. Hasil kompilasi di tingkat fakultas ini kemudian diserahkan kepada Rektor untuk dibuat rekap ditingkat universitas.

Rektor mengkompilasi semua laporan dari tingkat fakultas dan membuat rekap laporan di tingkat universitas. Rektor bertanggung jawab dan berwenang untuk memverifikasi kebenaran laporan yang telah disahkan oleh Dekan. Untuk perguruan tinggi negeri maka laporan ini diserahkan atau dikirim langsung kepada Direktur Jenderal Pendidikan Tinggi setiap tahun. Laporan yang dikirim dalam bentuk *hardcopy* dan *softcopy*.

Pada perguruan tinggi yang diselenggarakan oleh masyarakat maka laporan diserahkan atau dikirim kepada Koordinator Perguruan Tinggi Swasta (Kopertis) untuk dikompilasi ditingkat Kopertis pada waktu yang telah ditetapkan. Kopertis bertanggung jawab dan berwenang untuk memverifikasi kebenaran laporan yang telah disahkan oleh Rektor perguruan tinggi.

Kopertis kemudian mengkompilasi dan membuat rekap semua perguruan tinggi yang menjadi tanggung jawabnya. Rekap laporan dalam bentuk *hardcopy* dan *softcopy*. Kopertis kemudian menyerahkan dan atau mengirimkan laporan ke Direktur Jenderal Pendidikan Tinggi.



Gambar 2.9 Prosedur Kompilasi Database  
(Tim Dikti, 2010)

Kompilasi *database* evaluasi beban kerja dosen menampung data yang sangat banyak. Berdasarkan website dikti yang diakses pada 14 januari 2012 tercatat ada 3105 perguruan tinggi (**Tabel 2.5**) dan 271.196 dosen tersebar di seluruh Indonesia (**Tabel 2.6**). Informasi pada database evaluasi beban kerja dosen dalam pelaksanaan tridharma tidak akan banyak memberikan manfaat terhadap manajemen jika tidak diolah menjadi pengetahuan.

Tabel 2.5 Rekapitulasi Jumlah Perguruan Tinggi

	Total		Universitas		Institut		Sekolah Tinggi		Akademi		Politeknik	
	PT	PS	PT	PS	PT	PS	PT	PS	PT	PS	PT	PS
Negeri	83	4098	48	3414	6	353	2	12			27	319
Kopwil I	345	1215	39	577	3	26	132	335	155	197	16	80
Kopwil II	207	702	25	318	1	10	99	240	72	99	10	35
Kopwil III	325	1510	49	797	10	106	142	422	115	147	9	38
Kopwil IV	473	1904	50	791	6	49	240	725	147	219	30	120
Kopwil V	119	540	18	272	4	24	40	128	49	69	8	47
Kopwil VI	245	982	34	524	2	18	76	233	112	142	21	65
Kopwil VII	325	1516	75	901	12	103	141	364	85	106	12	42
Kopwil VIII	135	526	28	279	3	25	70	165	29	41	5	16
Kopwil IX	349	1113	47	535	3	14	164	378	129	167	6	19
Kopwil X	245	818	21	307	2	32	117	319	98	126	7	34
Kopwil XI	163	457	18	165	1	3	77	189	56	65	11	35
Kopwil XII	91	363	14	175	2	13	51	121	17	28	7	26
Total	3105	15744	466	9055	55	776	1351	3631	1064	1406	169	876

(Sumber: <http://evaluasi.or.id/recap-total.php#> diakses pada 14 Januari 2012)

Tabel 2.6. Rekapitulasi Jumlah Dosen

	Pria	Wanita	Total
PTN	101.375	51.547	152.922
Kopertis I	6.489	4.740	11.229
Kopertis II	4.055	3.195	7.250
Kopertis III	11.046	6.941	17.987
Kopertis IV	11.174	6.134	17.308
Kopertis V	3.696	2.461	6.157
Kopertis VI	5.781	4.223	10.004
Kopertis VII	10.245	6.934	17.179
Kopertis VIII	3.696	1.788	5.484
Kopertis IX	5.311	3.433	8.744
Kopertis X	2.980	2.961	5.941
Kopertis XI	2.782	1.731	4.513
Kopertis XII	1.931	905	2.836
Total	173.153	98.043	271.196

(Sumber: <http://evaluasi.or.id/recap/recap-teacher-gender.php?flag=all> diakses pada 14 Januari 2012)

Fakta yang tampak pada institusi perguruan tinggi adalah bahwa pesatnya peningkatan jumlah edukasional data, tidak dibarengi dengan pemanfaatan data tersebut secara maksimal untuk peningkatan manajemen (Delavari, 2008). Diperlukan penggalian data (*data mining*) untuk memperoleh pengetahuan atau pengenalan pola agar sumber daya informasi lebih bermanfaat.

## 71.2. Landasan Teori

### 71.2.1. Data Mining

Istilah *data mining* memiliki beberapa padanan seperti *knowledge discovery* atau *pattern recognition*. Istilah *knowledge discovery* atau penemuan pengetahuan digunakan karena tujuan utama dari *data mining* memang untuk mendapatkan pengetahuan yang masih

tersembunyi di dalam bongkahan data. Istilah *pattern recognition* atau *pengenalan pola* pun tepat digunakan karena pengetahuan yang hendak digali memang berbentuk pola-pola yang mungkin juga masih perlu digali dari dalam bongkahan data yang tengah dihadapi. (Susanto, 2010).

*Data Mining* sebagai salah satu cabang ilmu yang relatif baru mempunyai potensi pengembangan yang sangat besar dan diprediksi akan menjadi salah satu yang paling revolusioner pada dekade ini (Larose, 2006). *Data Mining* sendiri merupakan sebuah proses ekstraksi informasi untuk menemukan pola (*pattern recognition*) yang penting pada tumpukan data dalam *database* sehingga menjadi pengetahuan (*knowledge discovery*). Fungsi-fungsi dalam *data mining* antara lain: fungsi deskripsi, fungsi estimasi, fungsi Prediksi, fungsi Klasifikasi, fungsi *Clustering* dan fungsi asosiasi (Larose, 2005).

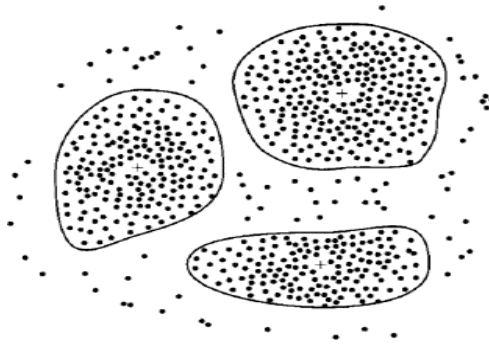
### 71.2.2. *Clustering*

*Clustering* adalah studi formal metode dan algoritma untuk partisi atau pengelompokan. Analisis *clustering* tidak menggunakan pelabelan kategori sebelumnya. *Clustering* bersifat *unsupervised learning* atau tidak mempunyai tahap pelatihan data, berbeda dengan klasifikasi (Jain, 2009).

*Clustering* adalah alat penemuan mengungkapkan hubungan dan struktur di dalam data yang sebelumnya tidak jelas menjadi pengetahuan yang bermanfaat ketika ditemukan.

Tujuan utama dari metode *clustering* adalah pengelompokan sejumlah data / obyek ke dalam *cluster (group)* sehingga dalam setiap *cluster* akan berisi data yang semirip mungkin. *Clustering* adalah metode yang berusaha untuk menempatkan obyek yang mirip (jaraknya dekat) dalam satu *cluster* dan membuat jarak antar *cluster* sejauh mungkin. Ini berarti obyek dalam satu *cluster* sangat mirip satu sama lain dan berbeda dengan obyek dalam *cluster-cluster* yang lain.

*Clustering* digunakan untuk menganalisis pengelompokan berbeda terhadap data, mirip dengan klasifikasi, namun pengelompokan belum didefinisikan sebelum dijalankannya *tool data mining*. *Clustering* membagi item menjadi kelompok-kelompok berdasarkan yang ditemukan *tool data mining*. Prinsip dari *clustering* adalah memaksimalkan kesamaan antar anggota satu kelas dan meminimumkan kesamaan antar *cluster*. *Clustering* dapat dilakukan pada data yang memiliki beberapa atribut yang dipetakan sebagai ruang multidimensi (Kusnawi, 2007). Ilustrasi dari *clustering* dapat dilihat pada **Gambar 2.10**. Singkatnya, *clustering* berusaha untuk menemukan komponen kelompok secara natural, berdasarkan pada kedekatan data.



Gambar 2.10 Ilustrasi Pengelompokan

### 71.2.3. Algoritma K-means

Konsep dasar dari algoritma K-means adalah pencarian pusat *cluster* (*centroid points*) secara iteratif. Pusat *cluster* ditetapkan berdasarkan jarak setiap data ke pusat *cluster*. Proses *clustering* dimulai dengan mengidentifikasi data yang akan di-*cluster*,  $x_{ij}$  ( $i=1,...,n; j=1,...,m$ ) dengan  $n$  adalah jumlah data yang akan di-*cluster* dan  $m$  adalah jumlah variabel.

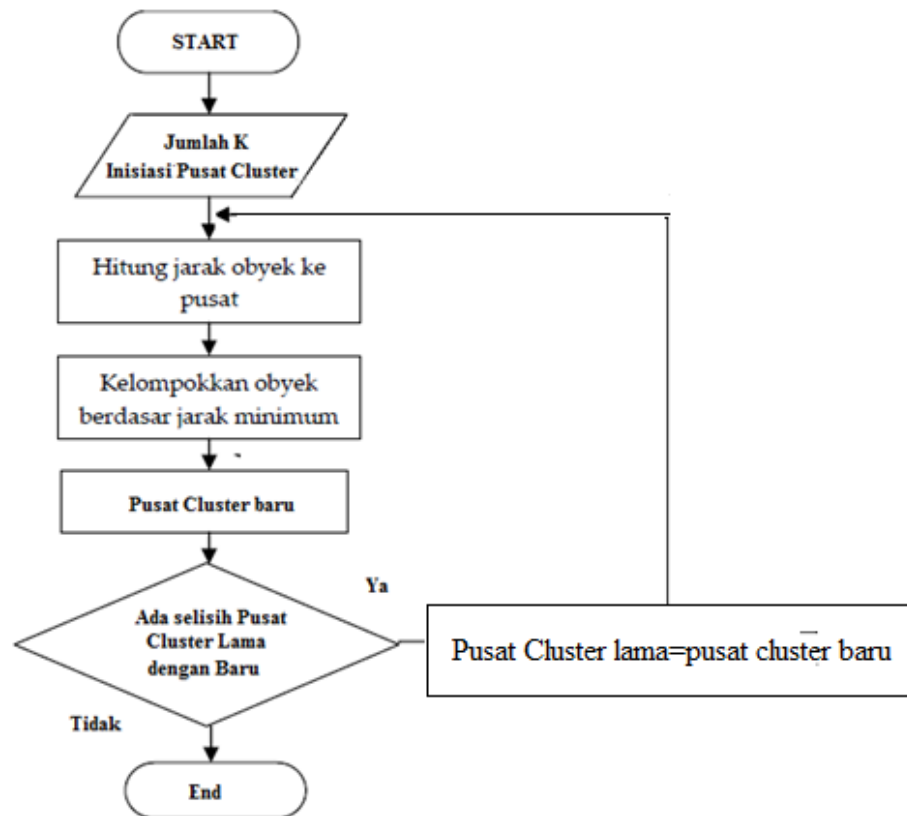
Pada awal iterasi, pusat setiap *cluster* ditetapkan secara bebas (sembarang),  $c_{kj}$  ( $k=1,...,k; j=1,...,m$ ). Kemudian dihitung jarak antara setiap data dengan setiap pusat *cluster*. Untuk melakukan penghitungan jarak data ke- $i$  ( $x_i$ ) pada pusat *cluster* ke- $k$  ( $c_k$ ), diberi nama ( $d_{ik}$ ), dapat digunakan formula Euclidean, yaitu:

$$d_{ik} = \sqrt{\sum_{j=1}^m (x_{ij} - c_{kj})^2} \quad (1)$$

Suatu data akan menjadi anggota dari *cluster* ke- $j$  apabila jarak data tersebut ke pusat *cluster* ke- $j$  bernilai paling kecil jika dibandingkan dengan jarak ke pusat *cluster* lainnya. Selanjutnya, kelompokkan data-data yang menjadi anggota pada setiap *cluster*. Nilai pusat *cluster* yang baru dapat dihitung dengan cara mencari nilai rata-rata dari data-data yang menjadi anggota pada *cluster* tersebut, dengan rumus:



$$c_{kj} = \frac{\sum_{h=1}^p y_{hj}}{p}; y_{hj} = x_{ij} \in \text{cluster ke-}k \quad (2)$$



Gambar 2.11 Flowchart Algoritma K-means

Algoritma K-means:

- Penetapan jumlah *cluster* (K), penetapan inisiasi pusat *cluster*
- Hitung jarak setiap data ke pusat *cluster* (**persamaan 1**)
- Kelompokkan data ke dalam *cluster* dengan jarak minimal
- Hitung pusat *cluster* baru berdasarkan rata-rata jarak terhadap pusat *cluster* (**persamaan 2**)
- Apakah ada selisih antara pusat cluster lama dengan pusat cluster baru? Jika ada, maka pusat cluster lama=pusat cluster baru.

- Ulangi langkah ii -iv hingga sudah tidak ada lagi selisih pada pusat *cluster*.

#### **71.2.4. Beban Kerja Dosen**

Dosen sebagaimana diamanatkan dalam UU Nomor 14 Tahun 2005 tentang Guru dan Dosen, dinyatakan sebagai pendidik profesional dan ilmuwan dengan tugas utama mentransformasikan, mengembangkan, dan menyebarkan ilmu pengetahuan, teknologi, dan seni melalui pendidikan, penelitian, dan pengabdian kepada masyarakat (Bab 1 Pasal 1 ayat 2).

Tugas utama dosen tersebut adalah melaksanakan tridharma perguruan tinggi dengan beban kerja paling sedikit sepadan dengan 12 (dua belas) sks dan paling banyak 16 (enam belas) sks pada setiap semester sesuai dengan kualifikasi akademiknya dengan ketentuan sebagai berikut:

- 1) Tugas melakukan pendidikan dan penelitian paling sedikit sepadan dengan 9 SKS yang dilaksanakan di perguruan tinggi yang bersangkutan
- 2) Tugas melakukan pengabdian kepada masyarakat dapat dilaksanakan melalui kegiatan pengabdian kepada masyarakat yang diselenggarakan oleh perguruan tinggi yang bersangkutan atau melalui lembaga lain sesuai dengan peraturan perundang undangan
- 3) Tugas penunjang tridharma perguruan tinggi dapat diperhitungkan sks nya sesuai dengan peraturan perundang undangan
- 4) Tugas melakukan pengabdian kepada masyarakat dan tugas penunjang paling sedikit sepadan dengan 3 SKS
- 5) Dosen yang mendapat penugasan sebagai pimpinan perguruan tinggi sampai dengan tingkat jurusan diwajibkan melaksanakan dharma pendidikan paling sedikit sepadan dengan 3 (tiga) sks
- 6) Dosen dengan status tugas belajar mempunyai tugas dan kewajiban belajar. Beban kerja dosen tugas belajar diatur dengan peraturan perundang undangan tersendiri (Tim Dikti, 2010).

### **BAB III**

#### **METODE PENELITIAN**

##### **72.1. Bahan Penelitian**

Bahan utama dari penelitian “*Data Clustering Menggunakan Metodologi CRISP-DM untuk Pengenalan Pola Proporsi Pelaksanaan Tridharma*” adalah:

- 1) Bidang Keilmuan: *Data Mining*
- 2) Metodologi / model : CRISP-DM
- 3) Teknik *data mining* : *Clustering*
- 4) Algoritma : K-means
- 5) Database : Beban Kerja Dosen dan Evaluasi Pelaksanaan Tridharma. Data simulasi menggunakan database Beban Kerja Dosen pada Universitas Lambung Mangkurat
- 6) Referensi: Pedoman Beban Kerja Dosen dan Evaluasi Pelaksanaan Tridharma Perguruan Tinggi yang diterbitkan Dikti tahun 2010.

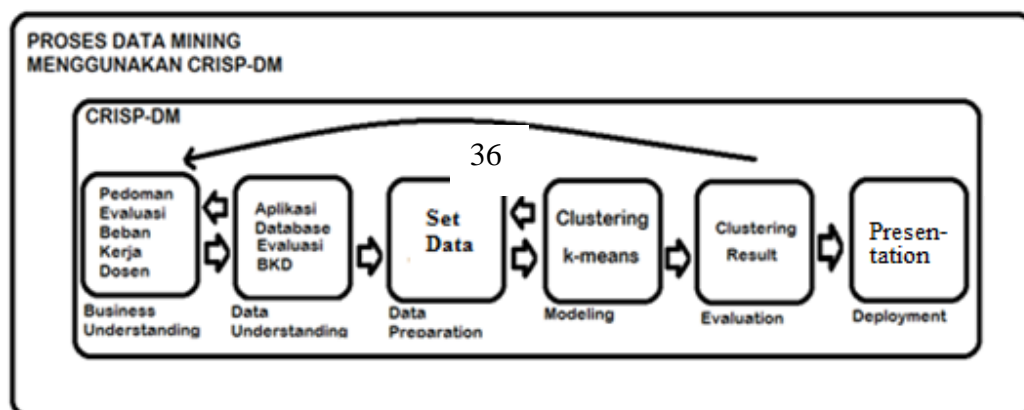
## 72.2. Alat Penelitian

Alat yang digunakan dalam penelitian ini adalah seperangkat komputer dengan spesifikasi standard. Kemudian perangkat lunak yang digunakan antara lain:

- 1) MS Access 2007, *database* dan aplikasi pemodelan data mining
- 2) Windows 7, sistem operasi.

## 72.3. Metode Penelitian

Metodologi *Cross Industry Standard Process for Data Mining* (CRISP-DM) digunakan sebagai standard proses *data mining* sekaligus sebagai metode penelitian (Gambar 3.1).



Gambar 3.1. Proses Data Mining berdasarkan Metodologi CRISP-DM

### 72.3.1. *Business understanding*

*Business understanding*, tahap pertama dalam proses CRISP-DM juga dapat disebut sebagai tahap pemahaman penelitian. Pada tahap ini dibutuhkan pemahaman tentang substansi dari kegiatan *data mining* yang akan dilakukan serta kebutuhan dari perspektif bisnis. Dalam rangka untuk memahami data yang kemudian dianalisis, sangat penting untuk sepenuhnya mempelajari objek bisnis (objek penelitian).

Pemahaman bisnis mengacu pada pedoman program Beban Kerja Dosen dan Evaluasi Pelaksanaan Tridharma yang diterbitkan Dikti pada tahun 2010. Pada tahap ini diperlukan pemahaman tentang latar belakang, tujuan, substansi dan semangat pada program tersebut. Pemahaman pada pedoman dan rubrik penjelasan program Beban Kerja Dosen ini diperlukan agar kemudian dapat mengubah pemahaman ini menjadi sebuah definisi masalah *data mining* dan mengembangkan rencana awal yang dirancang untuk mencapai tujuan.

#### 1) *Determine Business Objectives*

Tahap menentukan tujuan bisnis dan mengungkap faktor-faktor penting yang terlibat dalam penelitian yang direncanakan dan untuk memastikan bahwa penelitian tidak menghasilkan jawaban yang benar untuk pertanyaan yang salah.

Tujuan bisnis bukan merupakan tujuan langsung penelitian, namun lebih sebagai tujuan jangka panjang dalam dunia nyata. Hasil *data mining* mendeskripsikan sebuah pengetahuan yang akan menjadi tolak ukur institusi terkait dalam membuat kebijakan di masa akan datang.

Tujuan bisnis berdasarkan penjelasan pada pedoman program Beban Kerja Dosen antara lain untuk meningkatkan kinerja dosen dalam melaksanakan tridharma perguruan tinggi. Parameter suksesnya peningkatan kinerja dosen dalam pelaksanaan tridharma adalah dosen melaksanakan tridharma perguruan tinggi dengan beban kerja paling sedikit sepadan dengan 12 (dua belas) SKS dan paling banyak 16 (enam belas) SKS pada setiap semester dengan komposisi minimal 9 (sembilan) SKS untuk bidang pendidikan dan penelitian dan minimal 3 (tiga) SKS untuk bidang pengabdian masyarakat dan penunjang. Keseimbangan dalam pelaksanaan tridharma juga menjadi syarat keberhasilan dosen dalam tugas dan tanggungjawabnya.

#### 2) *Assess the Situation*

- a) Sistem sudah berjalan sesuai petunjuk pada pedoman evaluasi, namun masih diperlukan sosialisasi yang lebih mendalam
- b) Perguruan tinggi dan dosen masih beradaptasi dengan sistem pelaporan yang relatif baru ini
- c) Program aplikasi masih menggunakan sistem *stand alone* atau tidak terintegrasi dalam jaringan komputer
- d) Aturan validasi (*validation rules*) pada program aplikasi kurang baik, sehingga data tidak konsisten
- e) Hasil *data mining* terhadap database Beban Kerja Dosen tidak menggambarkan pola pelaksanaan tridharma dosen secara umum, namun hasil *data mining* menggambarkan pola proporsi tridharma pada periode pertama dan kedua (tahun ajaran 2009-2010 semester genap dan tahun ajaran 2010-2011 semester ganjil).

### 3) *Determine the Data Mining Goals*

Tahap mengubah pengetahuan pada domain bisnis menjadi sebuah definisi masalah *data mining* dan menentukan tujuan *data mining* (penelitian).

Tujuan data mining atau tujuan penelitian ini adalah menggali pengetahuan (*discovering knowledge*) tentang pola (*pattern*) proporsi pelaksanaan tridharma oleh dosen sehingga diketahui tingkat produktivitas, komposisi serta keseimbangan dalam pelaksanaan tridharma.

#### **72.3.2. Data Understanding**

*Data Understanding* atau pemahaman data adalah tahapan mengumpulkan data awal dan mempelajari data tersebut untuk bisa mengenal dan memahami apa saja yang bisa dilakukan pada data-data itu.

Pemahaman data mengacu pada database BKD (Beban Kerja Dosen). Tahap memahami format data secara permukaan (format *form* dan *report*) dan secara lebih mendalam (bentuk fisik data).

### 1) *Collect the Initial Data*

Tahap mengumpulkan data kompilasi hasil laporan dari *database* Beban Kerja Dosen dan Evaluasi Pelaksanaan Tridharma Perguruan Tinggi di lingkungan Universitas Lambung Mangkurat. *Database* yang digunakan pada penelitian ini adalah data BKD

periode 2009-2010 semester genap (periode pertama) dan periode 2010-2011 semester ganjil (periode kedua).

## 2) *Describe the Data*

Tahap memahami data yang tampak pada permukaan seperti laporan hasil, *form*, memeriksa format data, jumlah data, jumlah atribut-atribut data dan setiap fitur permukaan lain dari data.

Aplikasi BKD yang disebarakan Dikti ke semua perguruan tinggi negeri maupun swasta menggunakan *software* Microsoft Access. Data dan informasi yang tersedia bisa diperoleh dan dipelajari melalui *form* dan *report* nya saja disebabkan aplikasi tersebut masih berbentuk extension (.MDE).

Tahapan ini adalah tahap membiasakan diri dengan data-data yang sudah dikumpulkan dan berusaha menemukan wawasan awal mengenai informasi apa saja yang bisa didapatkan.

Data apa saja yang dimasukan pada aplikasi BKD dapat dipelajari dari form yang tersedia. Antara lain form Identitas Dosen (**Gambar 3.2**), form kinerja bidang pendidikan (**Gambar 3.3**), form kinerja bidang penelitian (**Gambar 3.4**), form kinerja bidang pengabdian pada masyarakat (**Gambar 3.5**), form kinerja bidang penunjang (**Gambar 3.6**), laporan evaluasi tingkat fakultas (**Gambar 3.7**) dan laporan evaluasi tingkat perguruan tinggi (**Gambar 3.8**). Tampilan *form* dan laporan tersebut dapat memberikan gambaran tentang data dan informasi apa yang bisa didapatkan dan dimanfaatkan untuk mencapai tujuan yang telah dirumuskan di tahap *business understanding*.

IDENTITAS DOSEN

**IDENTITAS DOSEN**

1. NO. SERTIFIKAT

2. NIP  NIDN

3. NAMA

4. NAMA PT

5. ALAMAT PT

6. FAKULTAS/DEPARTEMEN

7. PROG. STUDI

8. JAB. FUNGSIONAL  GOL

9. TGL. LAHIR  -  -  TEMPAT LAHIR

10. S1

11. S2

12. S3

13. JENIS

14. BIDANG ILMU

15. NO. HP

16. TAHUN LAPORAN  SEMESTER

17. ASESOR 1

18. ASESOR 2

Close

Gambar 3.2 Data Identitas Dosen

KINERJA BIDANG PENDIDIKAN

NO	JENIS KEGIATAN	BEBAN KERJA		MASA PENUGASAN	KINERJA		REKOMENDASI
		BUKTI PENUGASAN	SKS		BUKTI DOKUMEN	SKS	
1	Memberi kuliah Listrik Magnet 2	SK Dekan MIPA No: 023a/H8.1.28/PS/2010	2	Semester Genap TA 2009/2010	DPNA	2	SELESAI
2	Memberi kuliah Elektronika Komunikasi	SK Dekan MIPA No: 023a/H8.1.28/PS/2010	2	Semester Genap TA 2009/2010	DPNA	2	SELESAI
3	Memberi kuliah Instrumentasi Industri	SK Dekan MIPA No: 023a/H8.1.28/PS/2010	2	Semester Genap TA 2009/2010	DPNA	2	SELESAI
4	Memberi kuliah Pengantar Robotika	SK Dekan MIPA No: 023a/H8.1.28/PS/2010	2	Semester Genap TA 2009/2010	DPNA	2	SELESAI
5	Membimbing Tugas Akhir atas nama: Annisa, Ayu, Muliawati, S. Fahlevi	SK Dekan MIPA No: 018/H8.1.28/PS/2010	1	Semester Genap TA 2009/2010	SK	1	SELESAI
0						0	

Delete Close

Gambar 3.3 Data Kinerja Dosen Bidang Pendidikan

KINERJA BIDANG PENELITIAN DAN PENGEMBANGAN ILMU

NO	JENIS KEGIATAN	BEBAN KERJA		MASA PENUGASAN	KINERJA		REKOMENDASI
		BUKTI PENUGASAN	SKS		BUKTI DOKUMEN	SKS	
1	Penelitian Hibah Fundamental dengan judul: Modifikasi nanostruktur core	SK No: 007/SP2H/PP/DP2M/III/2010	4	2 semester	Laporan Kemajuan	2	SELESAI
0						0	

Gambar 3.4 Data Kinerja Dosen Bidang Penelitian

KINERJA BIDANG PENGABDIAN KEPADA MASYARAKAT

NO	JENIS KEGIATAN	BEBAN KERJA		MASA PENUGASAN	KINERJA		REKOMENDASI
		BUKTI PENUGASAN	SKS		BUKTI DOKUMEN	SKS	
1	Membuat karya pengabdian kepada masyarakat	SK.019/H8.1.28/PG/2010	1	11 maret 2010	laporan kegiatan	1	SELESAI
0						0	

Gambar 3.5 Data Kinerja Dosen Bidang Pengabdian Masyarakat

BIDANG PENUNJANG TRIDHARMA PERGURUAN TINGGI							
NO	JENIS KEGIATAN	BEBAN KERJA		MASA PENUGASAN	KINERJA		REKOMENDASI
		BUKTI PENUGASAN	SKS		BUKTI DOKUMEN	SKS	
1	Tim Teknologi Informasi dan Komunikasi FMIPA Unlam	SK Dekan No: 45b/H8.1.28/PS/2010	1	Tiap semester	SK	1	SELESAI
2	Pemakalah pada Seminar Nasional Nanoteknologi ke 7 dengan judul: Size	Sertifikat	1	Semester Genap TA 2009/2010	Sertifikat	1	SELESAI
3	Pembimbing akademik untuk 29 mahasiswa	SK Dekan MIPA No: 012/H8.1.28/PS/2010	1	Semester Genap TA 2009/2010	SK	1	SELESAI
*	0					0	

Gambar 3.6 Data Kinerja Dosen Bidang Penunjang

**LAMPIRAN II REKAP FAKULTAS**

**LAPORAN EVALUASI TINGKAT FAKULTAS TAHUN 2010**

**FAKULTAS**  
**UNIVERSITAS**

: Matematika dan Ilmu Pengetahuan Alam  
 : Universitas Lambung Mangkurat

NO	No. Sertifikat	Nama Dosen	Semester Gasal				Semester Genap				Keahlian Khusus Profesor	Status	Kesimpulan
			PD	PL	PG	PK	PD	PL	PG	PK			
1	1982 0208 2005 011003	MUHAMMAD AHSAR KARIM	7,5	4	1	3	0	0	0	0	0	DS	M
2	1983 0328 2005 012001	DEWI ANGGRAINI	9,5	1	1	3	0	0	0	0	0	DS	M
3	1007 10449	DEWI SRI SUSANTI	8,8	1	1	4	0	0	0	0	0	DS	M
4	0810 1009 161	FAISAL	8	1	1	4	0	0	0	0	0	DS	M
5	1979 1122 2008 012013	NA'IMAH HURATI	9	2	1	4	0	0	0	0	0	DS	M
6	1977 0813 2005 011003	NUR SALAM	8,8	1	1	4	0	0	0	0	0	DS	M
7	1007 10448	THRES YE, S.Si, M.Si	9	2	1	2	0	0	0	0	0	DS	M
8	1981 1010 2006 012004	YUNI YULIDA	8	3	1	2	0	0	0	0	0	DS	M
9	1976 0304 2001 121003	AHMAD BUDI JUNAIDI	5,5	4,2	1	4	0	0	0	0	0	DT	M
10	1011 0101 0396	AZDI IRWAN, S.Si, M.Si	9,5	0	0	5	0	0	0	0	0	DT	M
11	1982 1211 2005 01 2 001	DAHLENA ARIYANI, S.Si	10	3	0	3	0	0	0	0	0	DS	M
12	1981 0516 2008 01 2 023	DWI RASY MUJIYANTI	10	2	1	2	0	0	0	0	0	DS	M
13	0911 0100 8365	Edi Muktianto, S.Si, M.Si	8	3	3	1	0	0	0	0	0	DS	M
14	1976 1218 2000 122002	KHOLIFATU ROSYIDAH	6,5	4,8	1	2,3	0	0	0	0	0	DS	M
15	1978 0517 2001 122001	MARIA DEWI ASTUTI	8	1,6	1	4,3	0	0	0	0	0	DS	M
16	1011 0101 0397	Noer Komari	7	2	2	4	0	0	0	0	0	DT	M
17	1011 0101 0398	RADNA NURMASARI, S.Si, M.Si	5,3	5,4	1	4	0	0	0	0	0	DS	M
18	0810 1009 160	Drs. Rahmat Yunus, M.S.	8	1	1	6	0	0	0	0	0	DS	M
19	1977 0820 2005 011006	Sunardi, S.Si, M.Sc	8,3	3	2	2	0	0	0	0	0	DS	M
20	1011 0101 0395	UMI BAROROH LILI UTAMI, S.Si, M.Si	8,5	3,2	0	3	0	0	0	0	0	DS	M
21	1981 0214 2005 01 2 002	UTAMI IRAWATI	9	3	2	1	0	0	0	0	0	DT	M
22	1981 0919 2005 011004	Aditya Rahman K.N.S. Si, M.Eng	8	2	1	1	0	0	0	0	0	DT	M
23	1982 0228 2005 012001	Anni Nurliani, S.Si, M.Sc	9,6	3	1	2	0	0	0	0	0	DS	M
24	101-101-011-130148	Dr. Ir. Badruzsaufari, M.Sc	7,3	2,7	0	6	0	0	0	0	0	DS	M

Gambar 3.7 Laporan Evaluasi Tingkat Fakultas



### LAMPIRAN III REKAP UNIVERSITAS

#### LAPORAN EVALUASI TINGKAT UNIVERSITAS TAHUN 2009

NAMA UNIVERSITAS : UNIVERSITAS LAMBUNG MANGKURAT  
ALAMAT UNIVERSITAS : JL. BRIGJEND. H. HASAN BASRY BANJARMASIN 70123

NO	No. Sertifikat	Nama Dosen	Fakultas	Semester Gasal				Semester Genap				Kewajiban Khusus Profesor	Status	K
				PD	PL	PG	PK	PD	PL	PG	PK			
760	198209072008012009	NURLELY	FAKULTAS MATEMATIKA D	0	0	0	0	7	1	0	6	0	DS	
761	198212232008011004	LILING TRIYAS MONO	FAKULTAS MATEMATIKA D	0	0	0	0	9	0	1	5	0	DT	
762	197702142005011001	M Idris, S.Si	FAKULTAS MATEMATIKA D	0	0	0	0	12	0	0	0	0	DT	
763	198302022008041002	Muhammad Mafudh Shiddiq, S.Si	FAKULTAS MATEMATIKA D	0	0	0	0	12	0	0	0	0	DT	
764	197805172001122001	MARIA DEWI ASTUTI	FAKULTAS MATEMATIKA D	0	0	0	0	8,0	1	1	2	0	DS	
765	197408182002121002	Muhamat	FAKULTAS MATEMATIKA D	0	0	0	0	8	4	1	3	0	DS	
766	197911222008012013	NAIMAH HIRIATI	FAKULTAS MATEMATIKA D	0	0	0	0	9	1	1	4	0	DS	
767	198311152008121003	NASHRUL WATHAN	FAKULTAS MATEMATIKA D	0	0	0	0	8	1	2	1	0	DS	
768	08101009158	Dra. Nini Hadi Haryanti, M.S	FAKULTAS MATEMATIKA D	0	0	0	0	6	2	1	7	0	DT	
769	101101010397	Noer Komari	FAKULTAS MATEMATIKA D	0	0	0	0	7	2	2	4	0	DT	
770	198002192008012011	Noor Cahaya, S. Si, Apt	FAKULTAS MATEMATIKA D	0	0	0	0	12	0	0	0	0	DT	
771	197708132005011003	NUR SALAM	FAKULTAS MATEMATIKA D	0	0	0	0	9,5	1	1	3	0	DS	
772	197804142003122001	NURLINA, S. Si, M. So	FAKULTAS MATEMATIKA D	0	0	0	0	10	1	0	3	0	DS	
773	101101010392	Nurma Sari, S. Si, M. Si	FAKULTAS MATEMATIKA D	0	0	0	0	7,5	2	1	5	0	DS	
774	198104222008041003	NURUL HUDA, S.Si	FAKULTAS MATEMATIKA D	0	0	0	0	12	1	2	3	0	DS	
775	197301282005011003	Oni Soesanto, S. Si	FAKULTAS MATEMATIKA D	0	0	0	0	12	0	0	0	0	DT	
776	197808112005011001	Pardi Afandi, S. Si	FAKULTAS MATEMATIKA D	0	0	0	0	12	0	0	0	0	DT	
777	197910082003122001	Prihatin Oktiasari, S. Si, M. Si	FAKULTAS MATEMATIKA D	0	0	0	0	8	3	1	1	0	DT	
778	198212212008042002	Prima Happy, S. Farm. Apt	FAKULTAS MATEMATIKA D	0	0	0	0	12	0	0	0	0	DT	
779	198212042008011006	Radiyo Adi Nugroho, ST, M. Kom	FAKULTAS MATEMATIKA D	0	0	0	0	7	3	0	5	0	DS	
780	197303191997022001	RADNA NURMASARI, S. Si, M. Si	FAKULTAS MATEMATIKA D	0	0	0	0	5,8	5,2	1	4	0	DT	
781	08101009180	Drs. Rahmat Yunus M. Si	FAKULTAS MATEMATIKA D	0	0	0	0	8,5	8	2	10	0	DT	
782	001-010-19781220-200812-4004	M. Reza Faisal, ST	FAKULTAS MATEMATIKA D	0	0	0	0	8	1	0	0	0	DT	

Gambar 3.8 Laporan Evaluasi Tingkat Universitas

### 3) Explore the Data

Tugas ini melibatkan eksplorasi terhadap isi *database*. Eksplorasi data pada aplikasi Beban Kerja Dosen yang berbentuk *extention* (MDE) terbatas hanya pada permukaan data (*form* dan *report*). Penelusuran data hingga ke dalam *database* dapat dilakukan dengan cara mentransfer *tabel* pada *database* MDE ke dalam *database* ekstensi MDB terlebih dahulu.

Tabel 3.1 Tabel Mdos1 (Master Dosen)

Nama Field	Type Data	Keterangan
NoSerti *	Text	Nomor sertifikat sbg Primary Key
NIPDos	Text	NIP
NIDN	Text	NIDN
NMDos	Text	Nama Dosen
NMPT	Text	Nama PT Dosen
ALMPT	Text	Alamat PT Dosen
FakDos	Text	Fakultas Dosen
ProdiDos	Text	Prodi Dosen
Pangkat	Text	Pangkat
Gol	Text	Golongan
Tglhr	Date-Time	Tanggal Lahir Dosen
Tplhr	Text	Tempat Lahir Dosen
PTS1	Text	PT S1
PTS2	Text	PT S2
PTS3	Text	PT S3
KatDos	Text	Kategori Dosen (DS/DT)
BidIlDos	Text	Bidang Ilmu Dosen
NoHP	Text	No HP
KDPT	Text	Kode PT

SMTPT	Text	Tahun Ajaran
SEM	Number	Semester
NoAss1	Text	No Asesor 1
NMAss1	Text	Nama Asesor 1
NoAss2	Text	No Asesor 2
NMAss2	Text	Nama Asesor 2
KDKop	Text	Kode Kopertis
NMKop	Text	Nama Kopertis

Tabel 3.2 Tabel Trajar1 (Tridharma Pendidikan)

<b>Nama Field</b>	<b>Tipe Data</b>	<b>Keterangan</b>
SMTPT	Text	Tahun Ajaran + Semester
KDPT	Text	Kode PT
NoSerti	Text	No Sertifikasi (Foreign Key)
NoPend	Number	No Urut Kegiatan Pendidikan
NmPend	Text	Nama Kegiatan Pendidikan
BuktiPend	Text	Bukti Penugasan Kegiatan Pendidikan
SksPend	Number	SKS dibebankan
MasaPend	Text	Masa Penugasan Pendidikan
DokPend	Text	Dokumen Pelaksanaan Pendidikan
SksPend2	Number	SKS terlaksana
RekPend	Text	Rekomendasi Asesor

Tabel 3.3 Tabel Trpen1 (Tridharma Penelitian)

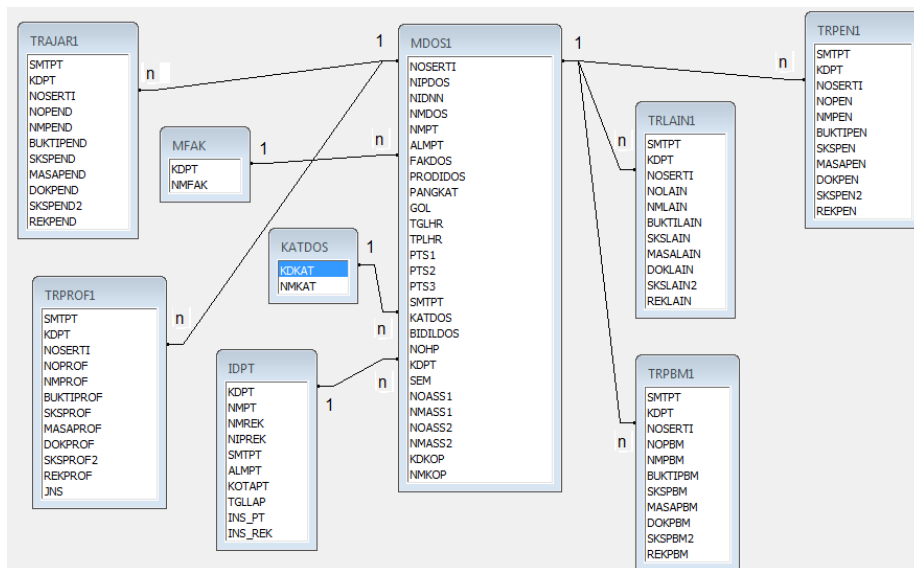
<b>Nama Field</b>	<b>Tipe Data</b>	<b>Keterangan</b>
SMTPT	Text	Tahun Ajaran + Semester
KDPT	Text	Kode PT
NoSerti	Text	No Sertifikasi (Foreign Key)
NoPen	Number	No Urut Kegiatan Penelitian
NmPen	Text	Nama Kegiatan Penelitian
BuktiPen	Text	Bukti Penugasan Kegiatan Penelitian
SksPen	Number	SKS dibebankan
MasaPen	Text	Masa Penugasan Penelitian
DokPen	Text	Dokumen Pelaksanaan Penelitian
SksPen	Number	SKS terlaksana
RekPen	Text	Rekomendasi Asesor

Tabel 3.4 Tabel Trpbm1 (Tridharma P2M)

<b>Nama Field</b>	<b>Tipe Data</b>	<b>Keterangan</b>
SMTPT	Text	Tahun Ajaran + Semester
KDPT	Text	Kode PT
NoSerti	Text	No Sertifikasi (Foreign Key)
NoP2M	Number	No Urut Kegiatan P2M
NmP2M	Text	Nama Kegiatan P2M
BuktiP2M	Text	Bukti Penugasan Kegiatan P2M
SksP2M	Number	SKS dibebankan
MasaP2M	Text	Masa Penugasan P2M
DokP2M	Text	Dokumen Pelaksanaan P2M
SksP2M2	Number	SKS terlaksana
RekP2M	Text	Rekomendasi Asesor

Tabel 3.5 Tabel Trlain1 (Tridharma Penunjang)

<b>Nama Field</b>	<b>Tipe Data</b>	<b>Keterangan</b>
NMTPT	Text	Tahun Ajaran + Semester
KDPT	Text	Kode PT
NoSerti	Text	No Sertifikasi (Foreign Key)
NoLain	Number	No Urut Kegiatan Penunjang
NmLain	Text	Nama Kegiatan Penunjang
BuktiLain	Text	Bukti Penugasan Kegiatan Penunjang
SksLain	Number	SKS dibebankan
MasaLain	Text	Masa Penugasan
DokLain	Text	Dokumen Pelaksanaan
SksLain2	Number	SKS terlaksana
RekLain	Text	Rekomendasi Asesor



Gambar 3.9 Relasi Antar Tabel

#### 4) *Verify Data Quality*

Tahap mengevaluasi kualitas data dan kelengkapan data. Nilai-nilai yang hilang sering terjadi, terutama jika data yang dikumpulkan di jangka waktu yang lama. Memeriksa atribut yang hilang atau kosong. Menilai apakah semua nilai masuk akal, ejaan nilai-nilai, dan apakah atribut dengan nilai yang berbeda memiliki arti yang sama.

Hasil penelusuran menemukan:

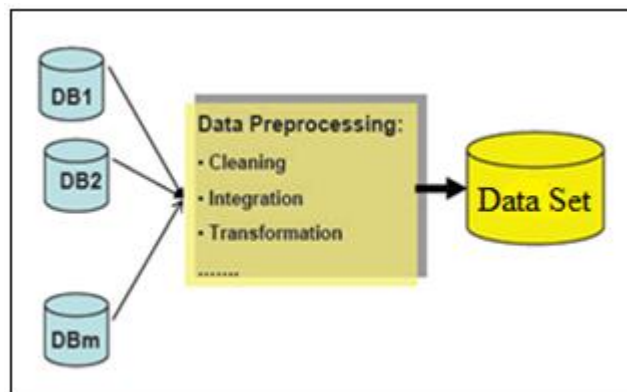
- Penetapan field NoSerti (No Sertifikasi Dosen) menjadi kunci. Padahal tidak semua dosen memilikinya. Meskipun diganti dengan NIP atau NIDN yang juga bersifat unik, namun akan menjadi masalah dalam hal konsistensi
- Format penulisan NIP Dosen yang tidak seragam. Ada yang menggunakan spasi, tanpa spasi dan tanda titik
- Dualisme pada master dosen. Setiap semester dosen selalu memasukan data idenditasnya sehingga sebagai master data hal ini membuat menjadi ambigu
- Atribut yang bernilai *null* atau kosong
- Format penulisan field lain seperti nama fakultas (FakDos), program studi (ProdiDos), pangkat dan golongan dosen yang tidak seragam.

NOSERTI	NIPDOS	NIDNN	NMDOS
0282/PTP-UNAIR/2008	19660808 199601 1 001	00-0808-6602	dr. Agung Biworo, M.Kes
1011010131501323	19732512 199903 1 001	0025127302	dr. H.Mohammad Bakhriansyah, M.Kes, Me
101102507356	197203071997021002	00-0703-7209	dr. Didik Dwi Sanyoto, M. Kes, M. Med.Ed
101102507353	197006151997021001	00070015067002	dr. Edyson, M. Kes.
101102507351	196401041994032001	0004016409	Dra.Fujiati,M.Si
19710415 199903 1 003	19710415 199903 1 003	0015047103	dr. Huldani
091101008362	19710627 199702 1 001	00-2706-7101	dr. Ahmad Husaini, M.Ag
091101008363	19710615 199702 2 002	00-1506-7103	dr. Lena Rosida, M.Kes
101102507350	1956053119870110	0031055602	dr.MiftahulArifin SpPK
197304231998032002	197304231998032002	00-2304-7302	dr. Noor Muthmainah, M.Sc
19770430 200501 2 003	19770430 200501 2 003	00-3004-7701	dr. Siti Wasilah, Msi.Med

Gambar 3.10 Format Data yang tidak Seragam

### 72.3.3. Data Preparation

*Data preparation* mencakup semua kegiatan untuk membangun *data set* yang akan dimasukkan ke dalam alat pemodelan dari data mentah awal atau membuat *database* baru untuk *set up data mining*. *Database* ini bersifat independen atau terpisah dari *database* operasional. Fungsi utamanya khusus untuk alat pemodelan *data clustering*. Persiapan data merupakan tahap yang padat karya. Tahap membangun *database* baru sebagai set data akhir untuk pemodelan *data mining clustering*.



Gambar 3.11 Ilustrasi Proses *Data Preparation*

#### 1) *Data Set Description*

Tahap merancang format *set data* mentah pada *database* parsial yang digunakan sebagai gudang data yang digunakan untuk pemodelan *data mining*. Rancangan set data harus menyesuaikan dengan apa yang telah dirumuskan pada tahapan *business understanding* terutama pada perumusan tujuan *data mining* yaitu mendeskripsikan pola (*pattern recognition*) proporsi pelaksanaan tridharma.

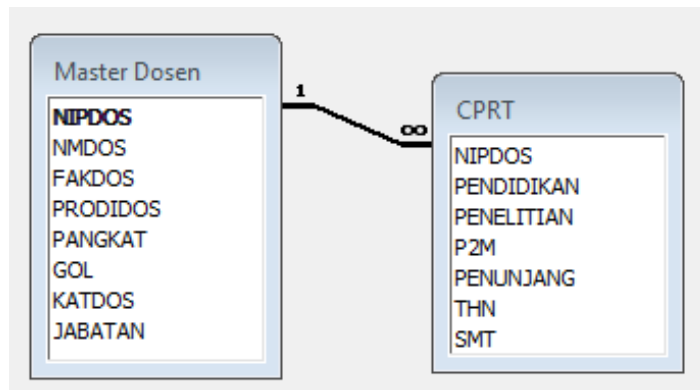
Tahap merancang format *set data* mentah pada *database* parsial yang digunakan sebagai gudang data yang digunakan untuk pemodelan *data mining*. Rancangan set data harus menyesuaikan dengan apa yang telah dirumuskan pada tahapan *business understanding* terutama pada perumusan tujuan *data mining* yaitu mendeskripsikan pola (*pattern recognition*) proporsi pelaksanaan tridharma. Pertama, perancangan untuk tabel rekapitulasi pelaksanaan tridharma dengan nama Tabel CPRT (**Tabel 3.6**). Kedua, perancangan untuk tabel master dosen baru yang lebih sederhana dengan nama Master Dosen. (**Tabel 3.7**). Fungsi Master Dosen untuk melihat hasil data clustering dari berbagai dimensi seperti usia dosen, masa kerja dan pangkat akademik.

Tabel 3.6 Rancangan Tabel Rekapitulasi Tridharma (CPRT)

No	Field	Tipe	Keterangan
1	NIPDos	Text	NIP Dosen
2	Pendidikan	Number	Jumlah SKS Bidang Pendidikan
3	Penelitian	Number	Jumlah SKS Bidang Penelitian
4	P2M	Number	Jumlah SKS Bidang Pengabdian Masyarakat
5	Penunjang	Number	Jumlah SKS Bidang Penunjang
6	THN	Text	Tahun Ajaran
7	SMT	Text	Semester

Tabel 3.7 Rancangan Tabel “Master Dosen”

No	Nama Field	Tipe Data	Keterangan
1	NIPDos *	Text	NIP sbg Primary Key
2	NMDos	Text	Nama Dosen
3	FakDos	Text	Fakultas Dosen
4	ProdiDos	Text	Prodi Dosen
5	Pangkat	Text	Pangkat
6	Gol	Text	Golongan
7	KatDos	Text	Kategori Dosen (DS/DT)
8	Jabatan	Text	Jabatan pada Tugas



Gambar 3.12 Relasi Antar Tabel untuk Database Baru

## 2) *Select Data*

Tahap memilih data yang digunakan untuk analisis didasarkan pada beberapa kriteria, termasuk relevansinya dengan tujuan *data mining*, serta kualitas dan kendala teknis seperti batas pada volume data atau jenis data.

Data yang dipilih pada setiap database adalah data yang berhubungan dengan pelaksanaan tridharma diambil dari tabel Trajar1, Trpen1, Trpbm1, Trlain1 dan tabel Mdos1 sebagai data master dosen.

## 3) *Construct Data*

Pengembangan set data baru atau memproduksi atribut turunan melalui proses transformasi data (*query*) yang sesuai dengan kebutuhan pemodelan. Transformasi melibatkan agregasi. Agregasi mengacu pada operasi dimana nilai-nilai baru dihitung dengan meringkas informasi dari beberapa catatan dan / atau tabel. Sebuah agregasi dapat mencakup konversi tabel, mencari total nilai, nilai rata-rata, persentase, dan lain-lain. Transformasi-transformasi ini mungkin diperlukan untuk mengubah rentang ke bidang simbolik. Pemodelan data mining atau algoritma sering membutuhkan transformasi-transformasi ini.

Setiap dosen dicari total SKS pelaksanaan setiap dharma-nya berdasarkan data yang dipilih pada tahap sebelumnya. Berikut ini adalah perintah memilih atribut dan menjumlahkan SKS setiap dharma dosen (**Gambar 3.13-3.16**). Kemudian dibuat rekapitulasi jumlah dharma setiap dosen dalam tabel baru bernama CPRT (**Gambar 3.17**).

SQL:SKS Trajar1

```
SELECT MDOS1.NOSERTI, Sum (TRAJAR1.SKSPEND2) AS SumOfSKSPEND2
```

```
FROM MDOS1 INNER JOIN TRAJAR1 ON MDOS1.NOSERTI = TRAJAR1.NOSERTI GROUP BY MDOS1.NOSERTI;
```

Field:	NOSERTI	SKSPEND2
Table:	MDOS1	TRAJAR1
Total:	Group By	Sum
Sort:		
Show:	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
Criteria:		

Gambar 3.13 Rekapitulasi SKS Bidang Pendidikan

### SQL:SKS Trpen1

```
SELECT MDOS1.NOSERTI, Sum(TRLAIN1.SKSLAIN2) AS SumOfSKSLAIN2
FROM MDOS1 INNER JOIN TRLAIN1 ON MDOS1.NOSERTI = TRLAIN1.NOSERTI GROUP BY MDOS1.NOSERTI;
```

Field:	NOSERTI	SKSPEN2
Table:	MDOS1	TRPEN1
Total:	Group By	Sum
Sort:		
Show:	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
Criteria:		

Gambar 3.14 Rekapitulasi SKS Bidang Penelitian

### SQL:SKS Trp2m1

```
SELECT MDOS1.NOSERTI, Sum(TRPBM1.SKSPBM2) AS SumOfSKSPBM2
FROM MDOS1 INNER JOIN TRPBM1 ON MDOS1.NOSERTI = TRPBM1.NOSERTI GROUP BY MDOS1.NOSERTI;
```

Field:	NOSERTI	SKSPBM2
Table:	MDOS1	TRPBM1
Total:	Group By	Sum
Sort:		
Show:	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
Criteria:		

Gambar 3.15 Rekapitulasi SKS Bidang P2M

### SQL:SKS Trlain1

```
SELECT MDOS1.NOSERTI, Sum(TRPEN1.SKSPEN2) AS SumOfSKSPEN2
```



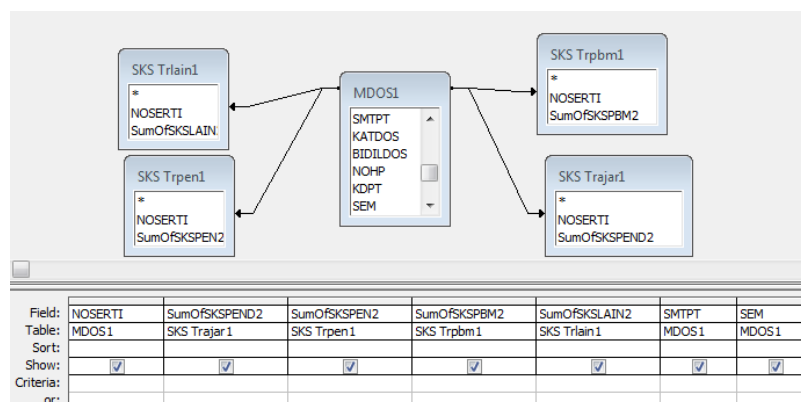
```
FROM MDOS1 INNER JOIN TRPEN1 ON MDOS1.NOSERTI = TRPEN1.NOSERTI GROUP BY MDOS1.NOSERTI;
```

Field:	NOSERTI	SKSLAIN2
Table:	MDOS1	TRLAIN1
Total:	Group By	Sum
Sort:		
Show:	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
Criteria:		

Gambar 3.16 Rekapitulasi SKS Bidang Penunjang

#### SQL:Rekap Tridharma

```
SELECT MDOS1.NOSERTI, [SKS Trajar1].SumOfSKSPEND2, [SKS Trpen1].SumOfSKSPEN2, [SKS Trpbm1].SumOfSKSPBM2, [SKS Trlain1].SumOfSKSLAIN2, MDOS1.SMTPT, MDOS1.SEM
FROM ((MDOS1 LEFT JOIN [SKS Trlain1] ON MDOS1.NOSERTI = [SKS Trlain1].NOSERTI) LEFT JOIN [SKS Trpbm1] ON MDOS1.NOSERTI = [SKS Trpbm1].NOSERTI) LEFT JOIN [SKS Trpen1] ON MDOS1.NOSERTI = [SKS Trpen1].NOSERTI) LEFT JOIN [SKS Trajar1] ON MDOS1.NOSERTI = [SKS Trajar1].NOSERTI;
```



Gambar 3.17 Query Rekapitulasi SKS Tridharma Dosen Persemester

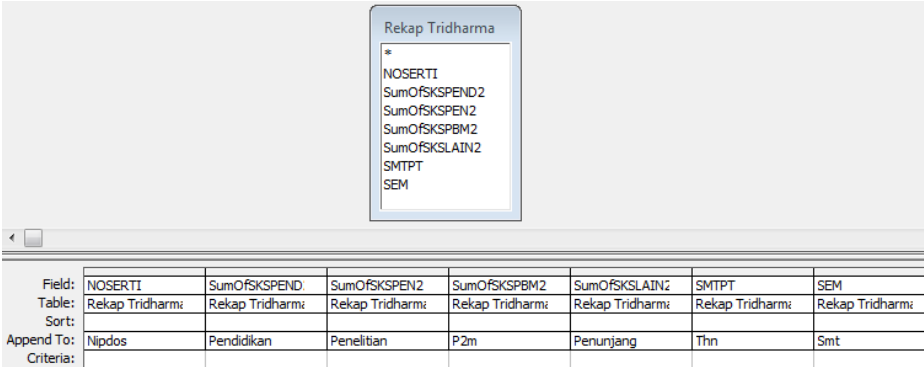
#### 4) Integrate Data

Tahap menggabungkan dua atau lebih tabel yang memiliki informasi yang berbeda tentang objek yang sama ke dalam set data baru yang telah disiapkan pada tahap awal *data preparation*. Tabel yang diintegrasikan berupa hasil transformasi data pada tahap sebelumnya.

Hasil rekapitulasi SKS tridharma setiap semester pada Query Rekap Tridharma diintegrasikan pada Tabel CPRT (**Tabel 3.6**).

## SQL:Integrate CPRT

```
INSERT INTO CPRT ( Nipdos, Pendidikan, Penelitian, P2m, Penunjang, Thn,
Smt )
SELECT [Rekap Tridharma].NOSERTI, [Rekap Tridharma].SumOfSKSPEND2, [Rekap
Tridharma].SumOfSKSPEN2, [Rekap Tridharma].SumOfSKSPBM2, [Rekap
Tridharma].SumOfSKSLAIN2, [Rekap Tridharma].SMTPT, [Rekap Tridharma].SEM
FROM [Rekap Tridharma];
```



Field:	NOSERTI	SumOfSKSPEND2	SumOfSKSPEN2	SumOfSKSPBM2	SumOfSKSLAIN2	SMTPT	SEM
Table:	Rekap Tridharma	Rekap Tridharma	Rekap Tridharma	Rekap Tridharma	Rekap Tridharma	Rekap Tridharma	Rekap Tridharma
Sort:							
Append To:	Nipdos	Pendidikan	Penelitian	P2m	Penunjang	Thn	Smt
Criteria:							

Gambar 3.18 Integrasi Rekapitulasi Tridharma

### 5) *Clean Data*

Tahap memastikan data yang dipilih, dikonstruksi dan diintegrasikan telah layak untuk ditambah. Kegiatannya antara lain membersihkan dan memperbaiki data yang rusak, menghapus data yang tidak di perlukan, menyeragamkan data yang dianggap sama namun memiliki nilai yang berbeda atau membuatnya menjadi konsisten.

Hasil verifikasi pada tahap *data understanding* menunjukkan adanya beberapa data dengan kualitas buruk. Data tersebut perlu dibersihkan dengan cara diperbaiki, dirubah atau dihapus. Data yang dibersihkan antara lain:

- NoSerti (No Sertifikasi Dosen) yang berfungsi sebagai primary key diganti dengan NIPDos (Nip Dosen). Alasannya adalah karena Nip Dosen dimiliki oleh semua dosen sedangkan Nomor Sertifikasi tidak. Ini semata untuk keseragaman
- Nip Dosen dilakukan penyeragaman format penulisan agar menjadi konsisten. Setiap karakter dalam Nip Dosen memiliki makna, yaitu tanggal lahir, tanggal mulai menjadi dosen (PNS) dan jenis kelamin

- c) Data master dosen pada setiap semester selalu diisi, hal ini menjadi memiliki dualisme data master dosen sehingga dipilih salah satu saja. Prinsip utamanya adalah bahwa data tersebut harus unik dan tidak ambigu atau memiliki dualisme. Maka data master dosen yang ganda akan dihapus salah satunya
- d) Pembersihan terhadap data yang tidak layak, seperti data tridharma yang null atau kosong dan jumlah SKS yang tidak wajar (*outlier*)
- e) Penyeragaman format penulisan pangkat, kategori dan fakultas dosen
- f) Penyesuaian nama field sesuai rancangan pada set data mentah (**Tabel 3.6 dan Tabel 3.7**).

Nipdos	Pendidikan	Penelitian	P2m	Penunjang	Thn	Smt
19660128 199303 2 002	10	2	1	3	2009	2
19800701 200812 2 001	10	2	1	2009	2	
101101003301	10	1	1	4	2009	2
197604142003122001	10	1		3	2009	2
19501115 198503 1 001	10	0,5	1	2	2009	2
091101008364	10	2		3	2009	2
8101006539	9,976	1,2		4	2009	2
101102507356	9,971	1,8		4	2009	2
091101011616	9,96	3		3	2009	2
101101004807	9,9	1	1	4	2009	2
0024036701	9,88			5,66	2009	2
101-101-004-615	9,85	3	1	2	2009	2
100710449	9,833	1	1	4	2009	2
198308202005012001	9,83	4	1	10	2009	2
09110100207	9,83	1,2	1	3	2009	2
091101011599	9,801		1	4,67	2009	2
19831226 200812 2 001	9,8	4		5,33	2009	2
101-101-004-733	9,75	0,5		3	2009	2

Gambar 3.19 Cuplikan Data Sebelum Pembersihan

NIPDOS	PENDIDIKAN	PENELITIAN	P2M	PENUNJANG	THN	SMT
196201081988111001	10	0	1	5	2009-2010	GENAP
196105101988031002	10	1	2	1	2009-2010	GENAP
195911261987102001	10	1	2	1	2009-2010	GENAP
196909291995021001	10	0	0	6	2009-2010	GENAP
198504142008122001	10	1	1	2	2009-2010	GENAP
195907101985031005	10	2	2	1	2009-2010	GENAP
196105031987032001	9,976	1,2	0	4	2009-2010	GENAP
197203071997021002	9,971	1,8	0	4	2009-2010	GENAP
195308071979031005	9,96	3	0	3	2009-2010	GENAP
196205301989032002	9,9	1	1	4	2009-2010	GENAP
196703241993031002	9,88	0	0	5,66	2009-2010	GENAP
195609011980031005	9,85	3	1	2	2009-2010	GENAP
197305161999032002	9,833	1	1	4	2009-2010	GENAP
196308211988031007	9,83	1,2	1	3	2009-2010	GENAP
195712131985031003	9,801	0	1	4,67	2009-2010	GENAP
198312262008122001	9,8	4	0	5,33	2009-2010	GENAP
196310281990032001	9,75	0,5	0	3	2009-2010	GENAP
197002091993031003	9,71	0	1	1,33	2009-2010	GENAP

Gambar 3.20 Cuplikan Data Setelah Pembersihan

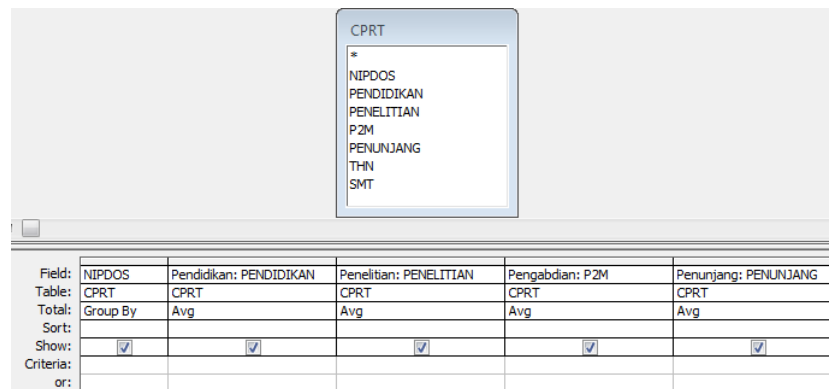
## 6) *Fotmat Data*

Tahap ini memproduksi *set data* akhir yang siap ditambangkan atau diolah dalam *tools* pemodelan *data mining*. Format set data akhir berupa sebuah *tabel / query (flat file)* untuk alat pemodelan data mining.

Data yang akan ditambang adalah SKS rata-rata dari pelaksanaan setiap tridharma dosen. Tabel CPRT yang berisikan data pelaksanaan tridharma dihitung SKS rata-ratanya dengan menggunakan query “Rata2 Tridharma”. Adapun proses pembuatannya sebagai berikut:

#### SQL:Rata2 Tridharma

```
SELECT CPRT.NIPDOS, Avg(CPRT.PENDIDIKAN) AS Pendidikan,
Avg(CPRT.PENELITIAN) AS Penelitian, Avg(CPRT.P2M) AS Pengabdian,
Avg(CPRT.PENUNJANG) AS Penunjang
FROM CPRT GROUP BY CPRT.NIPDOS;
```



Field:	NIPDOS	Pendidikan: PENDIDIKAN	Penelitian: PENELITIAN	Pengabdian: P2M	Penunjang: PENUNJANG
Table:	CPRT	CPRT	CPRT	CPRT	CPRT
Total:	Group By	Avg	Avg	Avg	Avg
Sort:					
Show:	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
Criteria:					
or:					

Gambar 3.21 Query “Rata2 Tridharma” untuk Set Data Pemodelan

#### 72.3.4. Modelling

*Modeling* adalah fase yang secara langsung melibatkan teknik data mining. Pemilihan teknik data mining, algoritma dan menentukan parameter dengan nilai yang optimal. Langkah-langkahnya pada *modelling* sebagai berikut:

##### 1) *Select Modelling Technique*

Teknik data mining yang dipilih adalah *clustering* dengan menggunakan algoritma K-means. *Clustering* dan algoritma K-means sangat tepat digunakan untuk mencapai tujuan awal penelitian ini yaitu menggali pengetahuan tentang pola pelaksanaan tridharma sekaligus untuk mengelompokkan dosen berdasarkan pola proporsi pelaksanaan tridharma.

Pemodelan *data mining* diawali dengan membangun aplikasi (*tools*) *clustering* set data akhir dengan algoritma K-means..

2) *Generate Test Design*

Tes pengujian atau tahap pembelajaran. Teknik *clustering* tidak memerlukan tahap pembelajaran kerana *clustering* bersifat *unsupervised learning* dan melakukan pengelompokan secara alamiah berdasarkan kemiripan atributnya, berbeda dengan teknik klasifikasi lainnya.

3) *Build Model*

*Software data mining* dibangun dengan menggunakan program aplikasi Microsoft Access 2007. Meskipun *tools data mining* yang menyediakan fitur algoritma K-means sudah cukup banyak, antara lain Rapid Miner, Weka, Matlab dan lain sebagainya, namun dengan membangun sendiri pemodelan *clustering* dengan algoritma K-means ini diharapkan dapat dilakukan eksplorasi terhadap proses iterasi maupun hasil akhir dari pemodelan tersebut.

Pembangunan aplikasi *clustering* menggunakan Microsoft Access tujuannya untuk menyesuaikan dengan database BKD, sehingga antara database sebagai sumber data dapat menjadi satu paket (tidak terpisah) dengan *tools* pemodelan *clustering* dosen berdasarkan proporsi pelaksanaan tridharma ini.

Algoritma aplikasi *clustering* menggunakan algoritma K-means seperti yang diilustrasikan pada flowchart K-Means (**Gambar 2.11**). Data set akhir yang dihasilkan pada tahap *data preparation* adalah objek yang diolah dalam pemodelan ini. Selain data set, matrix pusat *cluster* perlu tempat penyimpanan agar proses *clustering* dapat berjalan sebagaimana mestinya. Rancangan Tabel untuk menyimpan nilai pusat *cluster* lama dan baru adalah seperti pada **Tabel 3.8**.

Tabel 3.8 Rancangan Tabel Penyimpanan Pusat *Cluster*

No	Field Name	Tipe Data	Keterangan
----	------------	-----------	------------

1	Kn	Text	Nama Kluster
2	KP1	Number	Pusat Kluster Prosentase Bidang pendidikan
3	KP2	Number	Pusat Kluster Prosentase Bidang Penelitian
4	KP3	Number	Pusat Kluster Prosentase Bidang P2M
5	KP4	Number	Pusat Kluster Prosentase Bidang Penunjang

Jumlah *cluster* ( $k$ ) yang digunakan pada penelitian ini adalah 3 dengan pertimbangan bahwa *cluster* tersebut akan mewakili kelompok yang Bagus, Sedang dan Kurang. Meskipun metode *clustering* bersifat *unsupervised learning* dimana tidak ada pelabelan pada hasilnya, namun dengan menentukan jumlah  $k=3$ , diharapkan hasil *cluster* nanti mewakili kategori kelompok tersebut (bagus, sedang, kurang). Nama *cluster* ditentukan yaitu K1, K2 dan K3.

#### 4) *Assess Model*

Inisiasi pusat *cluster* yang berbeda dapat menghasilkan pusat *cluster* akhir yang berbeda pula. Meski pun inisiasi pusat *cluster* pada algoritma K-Means ditentukan secara acak, metode penentuan inisiasi diperlukan untuk mendapatkan hasil pusat *cluster* akhir yang optimal.

Pemodelan dilakukan beberapa kali dengan inisiasi pusat *cluster* yang berbeda agar dapat menilai pemodelan mana yang menghasilkan pusat *cluster* akhir yang paling optimal. Pengujian dilakukan dengan 11 pasang metode inisiasi pusat *cluster* yang berbeda.

Tabel 3.9 Inisiasi Pusat *Cluster*

NO	METODE INISIASI PUSAT <i>CLUSTER</i>
1	DATA COMPRESSED ATAS
2	DATA COMPRESSED 4,5,6
3	RANDOM MIN
4	RANDOM MAX
5	DATA COMPRESSED CHOSEN
6	DATA COMPRESSED BY BLOCK
7	K-MEANS BLOCK (3 BLOCKS)

8	K-MEANS CLOCK (10 BLOCKS)
9	K-MEANS BLOCK (10 BLOCKS) BY K-MEANS
10	STATISTIK TRIDHARMA
11	RAPIDMINER

Dari 11 percobaan itu akan diseleksi pusat *cluster* akhir yang paling optimal. Penentuan hasil *cluster* yang paling optimal dilakukan dengan cara mencari pusat *cluster* mayoritas dan memeriksa rasio antara besaran  $B(k)$ =*Between Cluster Variation* (BCV) dengan  $W(k)$ =*Within Cluster Variation* (WCV) seperti pada **Persamaan 3**. Rasio dengan nilai terbesar merupakan parameter hasil *cluster* terbaik (Stuart, 2004).

$$CH(k) = \frac{B(k)/(k-1)}{W(k)/(n-k)} \quad (3)$$

*Between Cluster Variation* (BCV) menyatakan jarak Euclides dari  $m_i$  ke  $m_j$  seperti pada **Persamaan 4** (Larose, 2005).

$$d_{ik} = \sqrt{\sum_{j=1}^m (x_{ij} - c_{kj})^2} \quad (4)$$

*Within Cluster Variation* (WCV) menyatakan *sum of squared errors* seperti pada **Persamaan 5** (Larose, 2005).

$$SSE = \sum_{i=1}^k \sum_{p \in C_i} d(p, m_i)^2 \quad (5)$$

#### 72.3.5. Evaluation

Evaluasi adalah fase interpretasi terhadap hasil *data mining*. Evaluasi dilakukan secara mendalam dengan tujuan agar hasil pada tahap modelling sesuai dengan sasaran yang ingin dicapai dalam tahap *business understanding*.

1) *Evaluate Results*

Tahap ini menilai sejauh mana hasil pemodelan *data mining* memenuhi tujuan data mining yang ditentukan pada tahap *business understanding*. Jika diaplikasikan dalam dunia nyata, tahap *evaluation* sebaiknya melibatkan pihak pengguna sistem atau yang kompeten misalnya pimpinan fakultas, pimpinan perguruan tinggi, pimpinan kopertis bahkan pimpinan Dikti.

2) *Review Process*

Tahap memeriksa kembali tahapan dari awal untuk memastikan bahwa tidak ada faktor penting dalam proses tersebut yang terabaikan atau terlewat.

3) *Determine Next Steps*

Tahap menentukan langkah apa yang diambil. Ada 2 pilihan: Kembali ke tahap awal (*business understanding*) atau melanjutkan ke tahap akhir (*deployment*).

### **72.3.6. Deployment**

*Deployment* merupakan tahapan membuat laporan hasil kegiatan data mining. Laporan akhir mengenai pengetahuan yang didapat atau pengenalan pola pada data dalam proses data mining dan dipresentasikan dalam bentuk grafik atau deskripsi yang mudah dipahami.